THE UNIVERSITY OF ALBERTA

ON SOME COMBINATORIAL AND PROBABILISTIC ASPECTS OF BIPARTITE GRAPHS

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

by

JOHN WESLEY MOON

EDMONTON, ALBERTA

ABSTRACT

The purpose of this thesis is to investigate various combinatorial and probabilistic properties of graphs when the restriction is imposed that the points of the graph have been separated into two distinguishable, mutually exclusive and exhaustive subsets and only points which belong to different subsets may be joined by an edge. Such graphs are often called bipartite graphs.

Expressions for the counting series for various types of such graphs are obtained in Chapter I.

Bipartite tournaments are considered in Chapter II. A few of their properties are developed and certain results known about ordinary tournaments are extended to the bipartite case. Ordinary tournaments often arise in connection with the method of paired comparisons and bipartite tournaments also provide natural models for certain situations. Some problems associated with some of these are treated.

In Chapter III the main emphasis is upon obtaining a few typical results of a structural nature for bipartite graphs. These can be interpreted as expressing some structural characteristic of a graph as a function of the number of points and edges it contains.

ACKNOWLEDGEMENTS

I wish to thank Professor Leo Moser, F.R.S.C., for his many helpful suggestions during the preparation of this thesis. Much of the work was done while I held a Province of Alberta Graduate Fellowship for which I would also like to express my appreciation.

TABLE OF CONTENTS

Page

CHAPTER I

ON SOME COUNTING PROBLEMS FOR BIGRAPHS

1.1 Definitions

A bipartite graph or, for the sake of brevity, a bigraph, consists of two finite distinguishable sets of things called points, P and Q, and a subset E of the set of all ordered pairs (p,q), where $p \in P$ and $q \in Q$. The elements of E are called edges and p and q are said to be adjacent, or joined, if, and only if, $(p,q) \in E$. The edge (p,q) is incident upon the points p and q. According to König [57], p. 170, the term "bipartite", or "bipartie", appears to have been used in this context first by Sainte-Laguë [88]. Two bigraphs with point sets P and Q, and P' and Q', respectively, are defined as being isomorphic if, and only if, there exists a one-to-one correspondence between the points of P and those of P', and between the points of Q and those of Q', respectively, which preserves all adjacency relations. The number of points in P and Q will be denoted by m and n, respectively. As trivial cases we shall admit the possibility that m or n equals zero but not both. By the degree of a given point in a given bigraph is meant the number of edges incident upon it. As general references on the theory of graphs see Berge [3] and König [57].

In this chapter we discuss how some of the methods used in enumerating ordinary graphs, i.e. graphs in which no distinction is made between points in P and points in Q as defined here, may be applied to the counting of bigraphs and obtain the counting series for various types of such graphs.

1.2 Pólya's theorem

Useful in many enumeration problems is a theorem due to Pólya [74]. Using the terminology of Harary [38] a statement of the theorem follows, restricted to the case of one variable, which will suffice for most of our purposes.

We take a figure to be an undefined term but assume that each figure may be characterized in some fashion by a single nonnegative integer called its content. If there are φ_k different figures of content k then the figure counting series, $\varphi(x)$, is defined by

$$\varphi(x) = \sum_{k=0}^{\infty} \varphi_k x^k \quad . \tag{1.2.1}$$

A configuration of length s is a sequence of s figures and its content is defined as being the sum of the contents of the figures of which it is composed. If G is some permutation group on s elements of order h then two configurations of length s are said to be G-equivalent if, and only if, there is a permutation of G, the configuration group, which takes one into the other. The configuration counting series, $F(x)$, is defined by

$$F(x) = \sum_{k=0}^{\infty} F_k x^k \quad , \tag{1.2.2}$$

where F_k denotes the number of G-inequivalent configurations of content k.

In order to express $F(x)$ in terms of G and $\varphi(x)$, the object

of the theorem, the cycle index of G, $Z(G)$, is needed. Let $h_{(j)}$ denote the number of permutations of G of type $(j) = (j_1, j_2, \ldots, j_s)$, i.e. having j_k cycles of length k, $k = 1, 2, \ldots, s$. Clearly

$$(1.2.3) \qquad 1 \cdot j_1 + 2 \cdot j_2 + \ldots + s \cdot j_s = s .$$

If $f_1, f_2, \ldots, f_s$ are indeterminates then the <u>cycle</u> <u>index</u> of G is defined by

$$(1.2.4) \qquad Z(G) = \frac{1}{h} \sum_{(j)} h_{(j)} f_1^{j_1} f_2^{j_2} \ldots f_s^{j_s} ,$$

where the sum is over all partitions (j) satisfying (1.2.3).

For example, the cycle index of S_n, the symmetric group of permutations of n elements, is (see e.g. Riordan [84], p. 68)

$$(1.2.5) \qquad Z(S_n) = \frac{1}{n!} \sum_{(j)} \frac{n!}{j_1! \, j_2! \ldots j_n!} \left(\frac{f_1}{1}\right)^{j_1} \left(\frac{f_2}{2}\right)^{j_2} \ldots \left(\frac{f_n}{n}\right)^{j_n} ,$$

summed over all partitions (j) satisfying (1.2.3) with s replaced by n.

For any function $f(x)$, let $Z(G; f(x))$ denote the function obtained from $Z(G)$ upon substituting $f(x^k)$ for f_k, $k = 1, 2, \ldots, s$. Then Pólya's theorem may be stated symbolically as follows:

$$(1.2.6) \qquad F(x) = Z(G; \varphi(x)) .$$

Some problems to be treated in a later section will require a slight modification of this theorem with a corresponding extension of the

definition of a cycle index to deal with the case where more than one kind of figure is involved.

1.3 The number of bigraphs

In this section we use Pólya's theorem to determine the number of nonisomorphic bigraphs with a given number of points and edges. Most of these results have been obtained by Harary [42] and are included here for the sake of completeness.

A bigraph may be regarded as a configuration whose figures are ordered pairs of points, the first point of each pair belonging to P and the second point of each pair belonging to Q. We define the content of such a figure to be one or zero according, respectively, as to whether the two points forming the pair are or are not joined by an edge in a given graph. Since these are the only possibilities in this case the figure counting series will be

$$(1.3.1) \qquad \varphi(x) = 1 + x .$$

The content of the graph, being the sum of the contents of its figures, will be equal to the total number of edges it contains and, hence, may range between 0 and mn for m by n bigraphs.

In determining if two m by n bigraphs are isomorphic any element of S_m, the symmetric group of permutations of m things, may be applied to the points in P, in conjunction with the application of any element of S_n to the points of Q. Hence the configuration group, under permutations of which some graphs will be considered equivalent, is isomorphic to $S_m \times S_n$, the direct product group of S_m and S_n.

However it is of degree mn since the combined effect of permuting points of P and Q among themselves, respectively, is the permuting of the mn ordered pairs of points, one from P and one from Q, among themselves.

Denote this configuration group by G_{mn} ; then upon setting

$$(1.3.2) \qquad g_{mn}(x) = \sum_{t=0}^{mn} g_{mn,t}\, x^t ,$$

where $g_{mn,t}$ denotes the number of nonisomorphic bigraphs with m points in P, n points in Q, and t edges, Pólya's theorem gives the result that

$$(1.3.3) \qquad g_{mn}(x) = Z(G_{mn} ; 1+x) .$$

It remains to obtain an expression for $Z(G_{mn})$. To accomplish this we first observe that formally multiplying $Z(S_m)$ by $Z(S_n)$, both of which may be regarded as known by (1.2.5), exhibits, with the correct coefficient, all the combinations of cycle types which can occur when permutations of S_m are applied to the points of P and permutations of S_n are applied to the points of Q. A typical term will have the appearance

$$\frac{h_{(j)}^{(m)}\, h_{(\ell)}^{(n)}}{m!\ \ n!}\, f_1^{j_1} f_2^{j_2} \cdots f_m^{j_m}\, t_1^{\ell_1} t_2^{\ell_2} \cdots t_n^{\ell_n} ,$$

where $h_{(j)}^{(m)}$ and $h_{(\ell)}^{(n)}$ denote the number of elements of S_m and S_n of type (j) and (ℓ), respectively, and where (j) and (ℓ) satisfy (1.2.3) with s replaced by m and n, respectively.

To determine the cycle character of the permutation on the figures of the graph induced by permutations of type (j) and (ℓ) being applied to the points of P and Q, respectively, we consider any figure whose first element lies in a cycle of length a, $1 \leq a \leq m$, and whose second element lies in a cycle of length b, $1 \leq b \leq n$. The P point of the figure first returns into itself after a applications of a given permutation of type (j) to the points of P and, similarly, the Q point of the figure first returns into itself after b applications of a given permutation of type (ℓ) to the points of Q. Hence, the complete figure is first returned into itself after $[a,b]$ simultaneous applications of permutations of type (j) and (ℓ) to the points of P and Q, respectively, where $[a,b]$ denotes the least common multiple of a and b.

Since there were j_a cycles of length a and ℓ_b cycles of length b in the respective permutations of the points of P and Q, it follows that there are $j_a \cdot a \cdot \ell_b \cdot b$ figures of the type we've just considered. As these are divided into cycles of length $[a,b]$ it follows that there are $j_a \ell_b (a,b)$ cycles of length $[a,b]$ among those figures whose points had the above properties, under the application of permutations of type (j) and (ℓ) to the points of P and Q, respectively, where (a,b) denotes the greatest common divisor of a and b.

The same argument applies to all such pairs, a and b, and yields the complete cycle characters of the permutation of the figures of the graph induced by the application of permutations of type (j)

and (ℓ) to the points of P and Q, respectively, since the points of every figure lie in a cycle of some length and each figure is counted exactly one.

Applying the same procedure to each of the terms in the product of $Z(S_m)$ and $Z(S_n)$ gives the result that

$$Z(G_{mn}) = \frac{1}{m!\,n!} \sum_{(j),(\ell)} h^{(m)}_{(j)}\, h^{(n)}_{(\ell)} \left[\prod_{a,b} f_{[a,b]}^{\,j_a \ell_b (a,b)} \right] , \tag{1.3.4}$$

where the sum is over the partitions, (j) and (ℓ), described above and the product is over pairs of integers, a and b, such that $1 \leq a \leq m$ and $1 \leq b \leq n$.

For example, one finds that

$$Z(G_{3,3}) = \frac{1}{36} (f_1^9 + 6f_1^3 f_2^3 + 9f_1 f_2^4 + 12f_3 f_6 + 8f_3^3) ,$$

and replacing f_i by $1+x^i$ gives the corresponding counting series for the number of nonisomorphic 3 by 3 bigraphs.

Using Pólya's theorem the first few nontrivial counting series for bigraphs are found to be

$$g_{2,2}(x) = 1 + x + 3x^2 + \ldots ,$$

$$g_{3,2}(x) = 1 + x + 3x^2 + 3x^3 + \ldots ,$$

$$g_{3,3}(x) = 1 + x + 3x^2 + 6x^3 + 7x^4 + \ldots ,$$

$$g_{4,2}(x) = 1 + x + 3x^2 + 3x^3 + 6x^4 + \ldots ,$$

$$(1.3.5)\qquad \begin{aligned} g_{4,3}(x) &= 1 + x + 3x^2 + 6x^3 + 11x^4 + 13x^5 + 17x^6 + \dots , \\ g_{4,4}(x) &= 1 + x + 3x^2 + 6x^3 + 16x^4 + 21x^5 + 39x^6 + 44x^7 \\ &\quad + 55x^8 + \dots , \\ g_{5,2}(x) &= 1 + x + 3x^2 + 3x^3 + 6x^4 + 6x^5 + \dots , \\ g_{5,3}(x) &= 1 + x + 3x^2 + 6x^3 + 11x^4 + 18x^5 + 26x^6 + 29x^7 + \dots , \\ g_{5,4}(x) &= 1 + x + 3x^2 + 6x^3 + 16x^4 + 27x^5 + 53x^6 + 80x^7 \\ &\quad + 120x^8 + 140x^9 + 159x^{10} + \dots , \\ g_{5,5}(x) &= 1 + x + 3x^2 + 6x^3 + 16x^4 + 34x^5 + 69x^6 + 130x^7 \\ &\quad + 234x^8 + 367x^9 + 527x^{10} + 634x^{11} + 755x^{12} + \dots , \end{aligned}$$

where the remaining terms in each series may be obtained by symmetry since there is a one-to-one correspondence between the nonisomorphic graphs with t edges and those with $mn-t$ edges induced by the process of <u>complementation</u>, i.e. forming a new graph in which two points, one of which is in P and the other is in Q, are joined by an edge if, and only if, they are not joined in the original graph.

Thus, for example, while there are $\binom{25}{12} = 5,200,300$ 5 by 5 bigraphs with 12 edges, the number of ways of selecting the 12 edges from the 25 that are possible, from (1.3.5) we see that these belong to only 755 essentially different types.

Following Harary [38] a bigraph of <u>strength</u> s may be defined as one in which there may be up to s indistinguishable edges joining a point in P to a point in Q; it is of <u>type</u> t if there are t distinguishable kinds of edges available for joining a point in P to

a point in Q where no more than one edge of any one kind joins two points. These two definitions may be combined to define a bigraph that is both of strength s and type t.

The definition of isomorphism of two bigraphs given previously may be extended in a straightforward manner to include these possibilities. The enumeration of the number of nonisomorphic graphs of these generalized types is accomplished in the same way as that of the ordinary ones with the exception that the figure counting series, for obvious reasons, are now

$$1 + x + \dots + x^s ,$$

(1.3.6)

$$(1 + x)^t ,$$

and

$$(1 + x + \dots + x^s)^t ,$$

respectively. Setting $s = t = 1$ gives (1.3.1) again.

A bigraph is said to be <u>directed</u> if for each edge (p,q) one of the points, p or q, is specified as being the <u>initial point</u> and the other point is specified as being the <u>terminal point</u> of the edge and it is further permitted that two directed edges may join p and q such that p is the initial point of one of them and q is the initial point of the other. Two directed bigraphs are isomorphic if they satisfy the previous definition of isomorphism with the additional requirement that the sense of direction of all edges is preserved in the correspondence.

By the <u>outdegree</u> of a point in a directed graph is meant the number of edges in the graph having it as the initial point. The <u>indegree</u> of a point has a similar definition.

The counting of nonisomorphic directed bigraphs proceeds along lines similar to the previously considered case. The only essential difference is that the configuration group is now of degree $2mn$ which results in the fact that the cycle index for the directed case can be obtained from that in (1.3.4) by increasing the exponent of each indeterminate by a factor of two. An <u>oriented</u> graph is, by definition, a directed graph in which not more than one oriented edge joins any pair of distinct points. It is not difficult to see (cf. Harary [40]) that the counting of nonisomorphic oriented bigraphs may be effected by substituting into the cycle index in (1.3.4) the figure counting series

$$1 + 2x \,. \tag{1.3.7}$$

Many other types of bigraphs may be counted in a similar fashion by performing appropriate modifications to the cycle index and the figure counting series. We may observe that for bigraphs there is, in a sense, a certain amount of overlap between the concepts of graphs of strength s and type t and directed and oriented graphs which does not carry through for more general types of graphs.

The expression given in (1.3.4) for the cycle index of G_{mn} is contained in Harary [42] as equation (7). In that paper when $m = n$ the sets P and Q are permitted to be interchanged in determining whether two bigraphs are isomorphic. To find the cycle index for this

case requires the use of a concept known as group exponentiation which we shall not discuss here. Also, instead of having just two sets of points such that only points in different sets may be joined, the more general situation where there may be an arbitrary number of such sets is mentioned.

Using an application of Pólya's theorem described in [38], which we shall use in §1.6, Harary and Prins [46] have obtained the counting series for such graphs which are connected, where by a connected graph is meant one such that for every ordered pair of distinct points, (a,b), there exists a path from a to b, i.e. a sequence of distinct points starting with a and ending with b such that successive points in the sequence are adjacent. They also have counted the number of nonisomorphic bicoloured graphs with no isolated points, where a bicoloured graph is an ordinary graph whose points have each been given one of two colours, but not both, such that each edge joins two points of different colour. An isolated point is a point of degree zero in the present context.

If $\bar{g}_{m,n}(x)$ denotes the counting series for nonisomorphic m by n bipartite graphs with no isolated points, then it is easily seen, for $m,n \geq 1$, that

$$(1.3.8) \qquad \bar{g}_{m,n}(x) = g_{m,n}(x) - g_{m-1,n}(x) - g_{m,n-1}(x) + g_{m-1,n-1}(x) .$$

Denoting the points of P and Q by $P_1, P_2, \ldots, P_m$ and $Q_1, Q_2, \ldots, Q_n$, respectively, one may say that two such labelled graphs are isomorphic if, and only if, they are identical. Lee [62]

has counted the number of labelled bipartite graphs with no isolated points.

Read [78] and Austin [2] have counted the number of labelled k-partite graphs, i.e. graphs with k distinct sets of labelled points such that no point is adjacent to another in the same set as itself. Wright [92] has investigated the asymptotic behavior of these numbers. Gilbert [30] has considered a related problem when $k = 2$ where, among other things, the edges are labelled as well and there is no restriction as to how many edges may join two points.

1.4 The number of rooted bipartite trees

By an m by n (bipartite) tree is meant an m by n bigraph which is connected and acyclic, i.e. contains no cycles or nontrivial paths whose initial and terminal points coincide. An m by n tree in which one P or Q point is singled out and distinguished from all the other points will be called a P or Q rooted tree, respectively. Two rooted trees are isomorphic if, and only if, they are isomorphic in the sense of the definition in §1.1 and, furthermore, their roots correspond.

Let $r_{mn}^{(P)}$ and $r_{mn}^{(Q)}$ denote the number of nonisomorphic m by n P and Q rooted trees, respectively. Then

$$(1.4.1) \qquad r_{1,0}^{(P)} = r_{0,1}^{(Q)} = 1 \quad \text{and} \quad r_{0,1}^{(P)} = r_{1,0}^{(Q)} = 0 \ .$$

In this section we indicate how the value of these numbers may be obtained.

Consider any m by n P rooted tree in which the root is of degree ℓ, where $1 \leq \ell \leq n$. Such a tree may be formed from $K_{0,1}$ 0 by 1 Q rooted trees, $K_{1,1}$ 1 by 1 Q rooted trees, $K_{1,2}$ 1 by 2 Q rooted trees, ... , $K_{j,h}$ j by h Q rooted trees, ... , and $K_{m-1,n}$ m-1 by n Q rooted trees, where

$$K_{0,1} + K_{1,1} + K_{1,2} + \dots + K_{j,h} + \dots + K_{m-1,n} = \ell ,$$

(1.4.2) $$0 \cdot K_{0,1} + 1 \cdot K_{1,1} + 1 \cdot K_{1,2} + \dots + j \cdot K_{j,h} + \dots + (m-1)\, K_{m-1,n} = m-1 ,$$

and

$$1 \cdot K_{0,1} + 1 \cdot K_{1,1} + 2 \cdot K_{1,2} + \dots + h \cdot K_{j,h} + \dots + n\, K_{m-1,n} = n ,$$

by joining each of the roots of the ℓ Q rooted trees to a new P point and calling this added point the root of the m by n tree thus constructed.

Counting the number of different ways in which this may be done gives

(1.4.3) $$r^{(P)}_{m,n} = \sum \prod_{\substack{j=0,\dots,m-1 \\ h=1,\dots,n}} \binom{r^{(Q)}_{j,h} + K_{j,h} - 1}{K_{j,h}} ,$$

where the sum is over $\ell = 1, 2, \dots, n$ and over all partitions, (K), satisfying (1.4.2), since the number of ways of selecting k things from a set of r with repetition permitted is $\binom{r+k-1}{k}$. (See, e.g. Riordan [84], p. 7.) As $r^{(P)}_{m,n} = r^{(Q)}_{n,m}$ this permits the

numbers to be computed recursively. Letting

$$r^{(P)}(x,y) = \sum_{\substack{m=1\\n=0}}^{\infty} r^{(P)}_{m,n} x^m y^n = r^{(Q)}(y,x)$$

(1.4.3) becomes

$$r^{(P)}(x,y) = x \prod_{\substack{j=0\\h=1}}^{\infty} (1 - x^j y^h)^{-r^{(Q)}_{j,h}} \quad . \tag{1.4.4}$$

If r_t denotes the number of nonisomorphic ordinary rooted trees on t points it is easily seen that

$$\sum_{m+n=t} r^{(P)}_{m,n} = r_t = \sum_{m+n=t} r^{(Q)}_{m,n} \quad .$$

Hence,

$$r(x) = \sum_{t=1}^{\infty} r_t x^t = r^{(P)}(x,x) = x \prod_{j=1}^{\infty} (1-x^j)^{-r_j} \quad , \tag{1.4.5}$$

which is a form due to Cayley [7].

Formally taking the logarithm of (1.4.4) gives

$$\begin{aligned} \log r^{(P)}(x,y) &= \log x + \sum_{\substack{j=0\\h=1}}^{\infty} r^{(Q)}_{j,h} \log(1-x^j y^h)^{-1} \\ &= \log x + \sum_{\substack{j=0\\h=1}}^{\infty} r^{(Q)}_{j,h} (x^j y^h + x^{2j} y^{2h}/2 + \ldots) \\ &= \log x + r^{(Q)}(x,y) + r^{(Q)}(x^2,y^2)/2 + \ldots \quad , \end{aligned}$$

or

$$r^{(P)}(x,y) = x \exp[r^{(Q)}(x,y) + r^{(Q)}(x^2,y^2)/2 + \dots] , \tag{1.4.6}$$

since these are only generating functions and convergence is immaterial.

Setting $x = y$ in this gives another analogous property of the generating function for the number of ordinary rooted trees.

Following Riordan [84], p. 133, in his treatment of the same problem for ordinary rooted trees, an alternate approach is to apply Pólya's theorem, where two variables are involved, to a configuration which consists of a P point, the root, joined to the roots of ℓ Q rooted bipartite trees. The figure counting series is simply $r^{(Q)}(x,y)$ and since the ℓ trees may be permuted in any fashion the cycle index is $Z(S_\ell)$. The root contributes an additional P point so the configuration counting series is, by Pólya's theorem, $x\,Z(S_\ell\,;\,r^{(Q)}(x,y))$. Summing over ℓ and using the identity (cf. Riordan [84], p. 68)

$$1 + \sum_{\ell=1}^{\infty} Z(S_\ell\,;\,f(x,y)) = \exp\left(\sum_{\ell=1}^{\infty} f(x^\ell,y^\ell)/\ell\right) , \tag{1.4.7}$$

with $f(x,y)$ replaced by $r^{(Q)}(x,y)$, one again obtains (1.4.6).

The first few terms of $r^{(P)}(x,y)$ are as follows:

$$\begin{aligned} r^{(P)}(x,y) = x &+ xy + xy^2 + xy^3 + xy^4 + xy^5 + \dots \\ &+ x^2y + 2x^2y^2 + 3x^2y^3 + 4x^2y^4 + 5x^2y^5 + \dots \\ &+ x^3y + 4x^3y^2 + 9x^3y^3 + 16x^3y^4 + 25x^3y^5 + \dots \\ &+ x^4y + 5x^4y^2 + 18x^4y^3 + 44x^4y^4 + 88x^4y^5 + \dots \\ &+ x^5y + 7x^5y^2 + 30x^5y^3 + 98x^5y^4 + 252x^5y^5 + \dots \ . \end{aligned} \tag{1.4.8}$$

1.5 A dissimilarity characteristic theorem for bipartite trees

The set of all transformations of a bigraph into itself which preserve all adjacency relations and the property of a points belonging to P or Q , respectively, form a group which is called the automorphism group of the graph. (See König [57], p. 5.) Two points of a graph are said to be similar if, and only if, there exists an element of the automorphism group of the graph which takes one point into the other. Similarity of edges is defined in the same way. The relation of being similar is an equivalence relation and separates the points and edges of a graph into mutually exclusive and exhaustive classes of similar points and edges, respectively. To facilitate the counting of the number of nonisomorphic bipartite trees in terms of the number of nonisomorphic rooted bipartite trees we prove a dissimilarity characteristic theorem for bipartite trees. The proof follows that given by Otter [71] for a corresponding result for ordinary trees with some simplification arising from the fact that no similar points are adjacent in a bipartite tree.

Theorem 1.5.1. The dissimilarity characteristic of any bipartite tree, the number of classes of dissimilar points minus the number of classes of dissimilar edges, equals one.

If T is some bipartite tree consider a subset of the points and edges of T which themselves form a subtree, T' , such that no two distinct points or edges of T' are similar to each other. Let a point q in T' be joined to a point p which is in T but not in T'. If there exists an element α of the automorphism group of T such that $(p)\alpha = p'$, where p' is in T' , then we may make the

following assertion:

(1.5.1) $(q)\alpha = q' = q$.

The removal of ℓ , the edge joining p and q , splits T into two bipartite trees, T_p and T_q , from the definition of a bipartite tree. Also, the removal of ℓ' , the edge joining p' and q' and which exists from the definition of α , splits T into two bipartite trees, $T_{p'}$ and $T_{q'}$. If $q' \neq q$, then ℓ' is not in T' since, by hypothesis, there were no distinct similar points in T'. Hence T' is included in $T_{p'}$, as is the edge ℓ. Clearly, p' , q' , and ℓ' are in T_q. Since ℓ is not in $T_{q'}$ it follows that $T_{q'}$ is included in T_q. Furthermore, p' is in T_q but not in $T_{q'}$. But, since T is a finite configuration, this contradicts the fact that $(T_q)\alpha = T_{q'}$. A simple diagram makes the reasons for some of the statements clearer.

Now let T' be the largest subtree of T , in terms of the total number of points involved, having no distinct similar points or edges. Then any point of T which is not in T' but which is joined to a point in T' must be similar to some point of T'. By induction it follows immediately that every point in T , not in T' , is similar to one, and of necessity only one, point in T'. Consider any edge, ℓ , which joins a point p to a point q and which is not in T'. Then there exists an element, α , of the automorphism group of T such that $(q)\alpha$ is in T'. If $(p)\alpha$ is in T' also then ℓ is similar to an edge of T'. If not, there exists another automorphism, β , such that $((p)\alpha)\beta$ is in T'. By (1.5.1) $((q)\alpha)\beta = (q)\alpha$, so

in any case we have that every edge of T not belonging to T' is similar to one, and only one, edge of T'. Hence, there are as many dissimilar classes of points and edges in T as there are points and edges, respectively, in T'. But T' is itself a tree and the number of points it contains minus the number of edges it contains equals one, by a form of Euler's theorem about polyhedra due to Listing [63]. (See also König [57], p. 51.) This suffices to prove Theorem 1.5.1.

The theorem may also be deduced as a special case of the dissimilarity characteristic theorem for Husimi trees or ordinary linear graphs contained in Harary and Norman [35], [36], and [45].

1.6 The number of bipartite trees

To determine $T_{m,n}$, the number of nonisomorphic m by n trees, we observe that the total number of dissimilar points minus the total number of dissimilar edges in all nonisomorphic m by n trees is $T_{m,n}$, by Theorem 1.5.1. The first quantity is simply the total number of rooted m by n trees and the second is the number of ways of forming an m by n tree by joining the root of a P rooted tree to the root of a Q rooted tree. Thus

$$T_{m,n} = r^{(P)}_{m,n} + r^{(Q)}_{m,n} - \sum_{\substack{j=1,\dots,m \\ h=0,\dots,n-1}} r^{(P)}_{j,h} \cdot r^{(Q)}_{m-j,n-h} \quad .$$

In the counting series form this becomes

$$(1.6.1) \qquad T(x,y) = \sum_{m,n=0}^{\infty} T_{m,n}\, x^m y^n = r^{(P)}(x,y) + r^{(Q)}(x,y) - r^{(P)}(x,y)\cdot r^{(Q)}(x,y),$$

with $T_{0,0} = 0$ by definition. Since the right member involves quantities

that may be regarded as known this enables the left member to be evaluated also. One finds the first few terms to be as follows:

$$(1.6.2)\qquad \begin{aligned} T(x,y) = x + xy + xy^2 &+ xy^3 + xy^4 + xy^5 + \ldots \\ &+ x^2y^2 + 2x^2y^3 + 2x^2y^4 + 3x^2y^5 + \ldots \\ &+ 4x^3y^3 + 7x^3y^4 + 10x^3y^5 + \ldots \\ &+ 14x^4y^4 + 28x^4y^5 + \ldots \\ &+ 65x^5y^5 + \ldots \quad . \end{aligned}$$

If $t(x,y)$ is the counting series for acyclic bigraphs, i.e. bigraphs each of whose connected components is a bipartite tree, then by applying Pólya's theorem to a configuration whose figures are bipartite trees and using the identity in (1.4.7) one has that

$$(1.6.3)\qquad 1 + t(x,y) = \exp\Big[\sum_{\ell=1}^{\infty} T(x^\ell,y^\ell)/\ell\Big] \quad .$$

A corresponding result may be written for acyclic bigraphs each of whose connected components is a rooted bipartite tree.

Riordan, in [83] and [85], and Harary and Prins [44] have counted ordinary trees in terms of several different parameters, the bipartite analogues of which could also be considered. Riordan, in the first paper above, counts the number of trees on n points whose points may be coloured in c, an arbitrary positive integer, colours such that no two points having the same colour are adjacent but this does not contain (1.6.1). Also, contained in these papers are references to methods of counting ordinary trees which do not make use of a dissimilarity

characteristic theorem and which could undoubtedly be adapted to the bipartite case, although they probably would not be any simpler than the method used here.

1.7 The number of labelled bipartite trees

The definition of a labelled P or Q rooted m by n tree combines the definitions of a labelled m by n bigraph with that of a P or Q rooted m by n tree, both of which have already been stated. In this section we derive formulae for $R_{m,n}^{(P)}$ and $R_{m,n}^{(Q)}$, the respective number of these, where

$$(1.7.1) \qquad R_{1,0}^{(P)} = R_{0,1}^{(Q)} = 1 \quad \text{and} \quad R_{0,1}^{(P)} = R_{1,0}^{(Q)} = 0 \quad .$$

Let $R_{m,n}^{(P)}(\ell)$ denote the number of labelled P rooted m by n trees in which the root is adjacent to precisely ℓ other points, where $1 \leq \ell \leq n$. Each such bipartite tree may be formed by combining ℓ labelled Q rooted trees by joining their roots to an added P point called the root, where the total number of P and Q points in the combined trees is m-1 and n, respectively. Upon counting the number of ways of forming these trees and labelling the points one finds that

$$(1.7.2) \qquad R_{m,n}^{(P)}(\ell) = \frac{m}{\ell!} \sum \frac{(m-1)!}{j_1! \ldots j_\ell!} \; \frac{n!}{h_1! \ldots h_\ell!} \; R_{j_1,h_1}^{(Q)} \cdots R_{j_\ell,h_\ell}^{(Q)} \quad ,$$

where the sum is over all partitions, (j) and (h), such that

$$j_1 + j_2 + \ldots + j_\ell = m-1 \quad ,$$

and

$$h_1 + h_2 + \ldots + h_\ell = n \quad .$$

The $\ell!$ factor arises from the fact that the ℓ edges joining the root to points which were previously roots may be permuted in any of $\ell!$ ways.

Then

$$(1.7.3) \qquad \sum_{m,n\geq 1}^{\infty} R_{m,n}^{(P)}(\ell) \frac{x^m y^n}{m!\, n!} = \frac{x}{\ell!} \left(\sum_{\substack{j=0\\h=1}}^{\infty} R_{j,h}^{(Q)} \frac{x^j y^h}{j!\, h!} \right)^{\ell} .$$

Denoting the sum in the parenthesis by $R^{(Q)}(x,y)$ and defining $R^{(P)}(x,y)$ similarly, summing over ℓ implies, since $R_{1,0}^{(P)} = 1$, that

$$(1.7.4) \qquad R^{(P)}(x,y) = x + x\, R^{(Q)}(x,y) + x[R^{(Q)}(x,y)]^2 / 2! + \ldots$$

$$= x \exp R^{(Q)}(x,y) \; .$$

$R^{(P)}(x,y) = R^{(Q)}(y,x)$ and Lagrange's formula may be used to determine the coefficients in these series. This is the approach used by Austin [2] and also by Scoins [90] who has recently rederived some of the results of Austin.

An alternate procedure consists in modifying a method used by Clarke [9] to derive the formula, due originally to Cayley [8], that there are n^{n-2} ordinary labelled trees on n points. Using Clarke's terminology we may proceed as follows.

In the present context rooted trees in which the root is of degree ℓ are said to be of type ℓ. An ordered pair of labelled

m by n P rooted trees, (X,Y), where X is of type $\ell-1$ and Y is of type ℓ, is called a linkage if, and only if, Y can be obtained from X by choosing one of the $n-(\ell-1)$ Q points not adjacent to the root and replacing the edge which goes from the point towards the root by an edge joining the point to the root directly. That such an edge exists and is unique follows from the fact that trees are connected and acyclic by definition. The total number of linkages for a given value of ℓ is then

$$(1.7.5) \qquad [n-(\ell-1)] \cdot R^{(P)}_{m,n}(\ell-1) .$$

The removal of the root of Y and the ℓ edges joining it to other points leaves ℓ distinct subtrees. Let there be t_i points of P in the ith such tree, for $i = 1, 2, \ldots, \ell$. Clearly $t_1 + t_2 + \ldots + t_\ell = m-1$. Now, all the bipartite trees of type $\ell-1$ linked to Y may be obtained by removing one of the ℓ edges incident upon the root of Y and replacing it by an edge joining the same Q point to a P point not in the same subtree as it is. Counting the number of ways of doing this gives another expression for the total number of linkages for a given value of ℓ, namely

$$(1.7.6) \qquad [(m-1-t_1) + (m-1-t_2) + \ldots + (m-1-t_\ell)]\, R^{(P)}_{m,n}(\ell) = (\ell-1)(m-1) R^{(P)}_{m,n}(\ell) .$$

Equating the expressions in (1.7.5) and (1.7.6) and using the rather obvious fact that $R^{(P)}_{m,n}(n) = mn^{m-1}$ implies the following result:

$$(1.7.7) \qquad R^{(P)}_{m,n}(\ell) = mn^{m-1} \binom{n-1}{\ell-1} (m-1)^{(n-1)-(\ell-1)}, \quad \ell = 1, \ldots, n .$$

For $m,n \geq 1$ every labelled P rooted m by n tree is of one of these types, so upon summing over ℓ we have that

$$(1.7.8) \qquad R^{(P)}_{m,n} = m\, m^{n-1}\, n^{m-1} .$$

An immediate consequence is that if $R_{m,n}$ denotes the number of labelled m by n trees, $m,n \geq 1$, then

$$(1.7.9) \qquad R_{m,n} = m^{n-1}\, n^{m-1} .$$

By other methods Austin [2] has counted what could be called labelled k-partite trees. His formula includes Cayley's result as a special case when each of the k sets of points contains only a single point. However there appear to be some inherent difficulties in trying to extend the approach used here to the more general situation.

Let $F_{m,n}(k)$ denote the number of labelled m by n bigraphs which consist of k disjoint bipartite trees such that $Q_1, Q_2, \ldots, Q_k$ all belong to separate trees, where $m \geq 1$ and $n \geq k \geq 1$. After deriving the solution to the analogous problem for ordinary graphs originally stated without proof by Cayley [8], Rényi [81] remarks that it can be used to derive the result corresponding to (1.7.7) given by Clarke [9] for ordinary graphs. The converse statement holds also and in our case we have

$$(1.7.10) \qquad R^{(P)}_{m+1,n}(k) = (m+1) \binom{n}{k} F_{m,n}(k) ,$$

since each graph counted in the left member may be obtained by choosing the label for the root in one of $m+1$ ways and then joining this root

to the points $Q_1, Q_2, \ldots, Q_k$, selected in 1 of $\binom{n}{k}$ ways, for each of the $F_{m,n}(k)$ graphs of the type described. With (1.7.7) this gives

$$(1.7.11) \qquad F_{m,n}(k) = k\, m^{n-k}\, n^{m-1} \; .$$

When $k = 1$ this gives (1.7.9).

The formula for the similar problem with respect to points in P may be obtained by symmetry.

To obtain a closed formula for the generalization of this problem to where points in both P and Q are specified as belonging to different trees appears to be more difficult although it is easy to write an expression for this in terms of the formulae just derived.

1.8 The number of connected bigraphs containing one cycle

After trees the simplest connected graphs are those which consist of a single cycle with trees attached to various points of the cycle. In this section we describe how to determine the number of bigraphs of this type.

Such a graph may be considered as a configuration whose figures are rooted trees whose root points are on a cycle of length $2k$, k an arbitrary integer ≥ 2, where by the length of a cycle is meant the number of edges it contains. The figure counting series is the counting series for rooted bipartite trees; but this depends upon whether the root point is in P or Q. We are thus presented with a situation in which we need to modify Pólya's theorem so that it will apply to a case where not only are figures characterized by their content but there are two distinguishable classes of figures to start with.

Let there be given two finite distinct sets of things of cardinality m and n, respectively, with not both m and n equal to zero. Let G be a subgroup of order h of the group $S_m \times S_n$ which permutes members of these two sets with members of the same set, respectively. We define a modified cycle index, $\bar{Z}(G)$, by

$$(1.8.1) \qquad \bar{Z}(G) = \frac{1}{h} \sum_{(j),(\ell)} h_{(j)(\ell)} f_1^{j_1} \cdots f_m^{j_m} t_1^{\ell_1} \cdots t_n^{\ell_n} ,$$

where $h_{(j)(\ell)}$ denotes the number of elements of G that are of type (j) and (ℓ) with respect to the elements of the first and second set, respectively. $f_1, \ldots, f_m$ and $t_1, \ldots, t_n$ are indeterminates and the summation is over the same partitions as in (1.3.4).

Now suppose the things in the first set to be figures of one kind and those in the second set to be figures of another kind. Defining equivalence with respect to G of two configurations each consisting of m figures of the first kind and n figures of the second kind in a way similar to the definition in §1.2 we may write, for reasons similar to those justifying (1.2.6), that

$$(1.8.2) \qquad \bar{F}(x) = \bar{Z}(G ; \varphi^{(1)}(x), \varphi^{(2)}(x)) ,$$

where $\bar{F}(x)$ denotes the configuration counting series and $\varphi^{(1)}(x)$ and $\varphi^{(2)}(x)$ denote the figure counting series for figures of the first and second kind, respectively. Corresponding to the ordinary case, the right member means the expression obtained by substituting $\varphi^{(1)}(x^i)$ for f_i, $i = 1, \ldots, m$ and $\varphi^{(2)}(x^k)$ for t_k, $k = 1, \ldots, n$, into the right member of (1.8.1).

(1.8.1) and (1.8.2) admit of further generalization to the case where figures are characterized with respect to more than one parameter - two in the problem to be treated here - and to where there are an arbitrary number of kinds of figures.

In the counting of the number of connected bigraphs with exactly one cycle the two figure counting series will be $r^{(P)}(x,y)$ and $r^{(Q)}(x,y)$, the counting series for P and Q rooted bipartite trees, respectively. We need to determine the modified cycle index determined by the symmetries of a cycle containing $k(\geq 2)$ P points and k Q points such that P points are taken only into P points and similarly for Q points.

For an ordinary cycle of length s in which all points are alike the group of symmetries is generated by a rotation,

$$R = (1 \ 2 \ldots s) ,$$

and by a reflection,

$$T = \begin{pmatrix} 1 & 2 & 3 & \ldots & s \\ 1 & s & s-1 & \ldots & 2 \end{pmatrix} ,$$

the numbers referring to the points. Riordan [84], p. 150, has derived the ordinary cycle index for this group, D_s, the dihedral group, and for $s = 2k$ it is

$$(1.8.3) \qquad Z(D_{2k}) = \frac{1}{4k}\left(\sum_{j|2k} \varphi(j)\, f_j^{2k/j} + k\, f_1^{\,2} f_2^{\,k-1} + k\, f_2^{\,k} \right),$$

where $\varphi(j)$ denotes the number of integers less than j and relatively

prime to j, with $\varphi(1) = 1$, and the sum is over the divisors of $2k$.

In cycles of the type we are considering all the even numbered points, starting the count at some point, say in P, belong to Q and all the odd numbered points belong to P. From the group of all symmetries of such a cycle we wish to retain only the subgroup of elements which take P points into P points and similarly with respect to Q points. Then we express the cycle behavior of these elements with respect to P and Q separately in the form of a modified cycle index.

The summed term in (1.8.3) corresponds to the rotations generated by the powers of the rotation denoted above by R. It is clear that the admissible transformations of this set will be those involving an even power of R. Since the cycle behavior of such transformations is the same with respect to P and Q the contribution of these to the corresponding modified cycle index will be

$$\sum_{d|k} \varphi(d)\, f_d^{k/d}\ \ t_d^{k/d}\ .$$

The term $k\, f_1^2\, f_2^{k-1}$ in the expression for $Z(D_{2k})$ arises from reflections typified above by T, which may be written as

$$T = (1)\,(2,2k)(3,2k-1)\ \ldots\ (k,k+2)(k+1)\ .$$

Suck reflections are all admissible in our case. Whether the point labelled $k+1$ is in the same set as the point labelled 1 depends on whether k is even or odd. Considering these two cases separately it is not difficult to see that the contribution of such reflections to the modified cycle index will be

$$\frac{k}{2}\left(f_1^{\,2}\, f_2^{\frac{k-2}{2}}\, t_2^{\frac{k}{2}} + f_2^{\frac{k}{2}}\, t_1^{\,2}\, t_2^{\frac{k-2}{2}}\right),$$

or

$$k\, f_1\, f_2^{\frac{k-1}{2}}\, t_1\, t_2^{\frac{k-1}{2}} \quad ,$$

according as to whether k is even or odd, respectively.

The term $k\, f_2^{\,k}$ arises from reflections typified by

$$T\, R = (1,2)\,(3,2k)\,\ldots\,(k,k+3)\,(k+1,k+2)\ .$$

All such transformations involve the interchanging of two points, the numbers associated with which are of different parity. Hence they will not be admitted to the group of symmetries which preserve the identity of the sets P and Q.

Combining the preceding expressions we have that $\bar{Z}(H_k)$, the modified cycle index determined by the group of symmetries of a cycle of length $2k$ formed by alternating points from two distinct sets which maintain the distinction between the two sets, may be written as

$$(1.8.4) \qquad \bar{Z}(H_k) = \begin{cases} \dfrac{1}{2k}\left(\displaystyle\sum_{d|k} \varphi(d)\, f_d^{k/d}\, t_d^{k/d} + \frac{k}{2}\, f_1^2\, f_2^{\frac{k-2}{2}}\, t_2^{\frac{k}{2}} + \frac{k}{2}\, f_2^{\frac{k}{2}}\, t_1^2\, t_2^{\frac{k-2}{2}}\right), \\[2ex] \dfrac{1}{2k}\left(\displaystyle\sum_{d|k} \varphi(d)\, f_d^{k/d}\, t_d^{k/d} + k\, f_1\, f_2^{\frac{k-1}{2}}\, t_1\, t_2^{\frac{k-1}{2}}\right), \end{cases}$$

according as to whether k is even or odd, respectively.

Letting $C_k(x,y)$ denote the counting series for the number of nonisomorphic connected bigraphs containing a single cycle, and that of length $2k$, $k \geq 2$, it follows from (1.8.2) that

$$(1.8.5) \qquad C_k(x,y) = \bar{Z}(H_k \, ; \, r^{(P)}(x,y) \, , \, r^{(Q)}(x,y)) \; .$$

The first few terms of $C_2(x,y)$, the simplest case, may be found to be as follows:

$$(1.8.6) \qquad \begin{aligned} C_2(x,y) = x^2y^2 + x^2y^3 &+ 2x^2y^4 + 2x^2y^5 + \ldots \\ + 3x^3y^3 &+ 7x^3y^4 + 10x^3y^5 + \ldots \\ &+ 19x^4y^4 + 43x^4y^5 + \ldots \\ &+ 138x^5y^5 + \ldots \; . \end{aligned}$$

When the points of the graph are labelled the corresponding problem has been solved by Austin [2].

1.9 The superposition theorem for bigraphs

Given k different ordinary graphs each on n points, the superposition theorem due to Read [76] may be used to determine, in terms of the automorphism groups of these graphs, the number of nonisomorphic graphs on n points which may be formed by identifying the points of the different graphs in a certain manner. The possibilities that more than one edge may join two points and that an edge may join a point to itself are admitted and using this theorem, and in conjunction with Pólya's theorem in a subsequent paper [77], Read counted several types of graphs, especially those in which certain restrictions are made on the number of edges that may be

incident on various points. In this section we derive the corresponding result arising from making the restriction that in all the graphs the points are partitioned into mutually exclusive and exhaustive classes in such a way that in the superpositioning process only points belonging to certain classes may be identified with each other. This will be defined more precisely in what follows.

Our argument, which parallels that of Read's and which retains his terminology insofar as possible, will be carried out only for the case where there are two graphs to be superposed and where the points in each have been split into two classes. However, it will be seen that the procedure is completely general.

Let there be given two finite sets of elements, P and Q, of cardinality m and n, respectively, where not both m and n are zero. Let H_i be a subgroup of order $h^{(i)}$, $i = 1, 2,$ of $S_m \times S_n$, the direct product group of the symmetric groups of degree m and n. Consider the set of all ordered pairs, (a_1, a_2), of elements of $S_m \times S_n$.

Two such ordered pairs, (a_1, a_2) and (b_1, b_2), will be said to be <u>L-similar</u> with respect to H_1 and H_2, or $(a_1, a_2)\ L\ (b_1, b_2)$, if, and only if, $b_i a_i^{-1} \in H_i$ for $i = 1, 2$. They will be said to be <u>T-similar</u> with respect to H_1 and H_2, or $(a_1, a_2)\ T\ (b_1, b_2)$, if, and only if, there exists an element x of $S_m \times S_n$ such that $(a_1 x, a_2 x)\ L\ (b_1, b_2)$. These both define equivalence relations and we wish to evaluate the number of equivalence classes, $(M{\cdot}N)_T$, determined by T.

An ordered pair is T-similar to (a_1, a_2) if, and only if, it is of the form $(h_1 a_1 x, h_2 a_2 x)$, $x \in S_m \times S_n$ and $h_i \in H_i$, for $i = 1, 2$. For fixed (a_1, a_2) there are $h^{(1)} h^{(2)} m!\, n!$ such pairs and we first determine how many times a typical one is repeated.

If $(g_1 a_1 y, g_2 a_2 y)$, $y \in S_m \times S_n$ and $g_i \in H_i$ for $i = 1, 2$, is the same ordered pair as $(h_1 a_1 x, h_2 a_2 x)$, then

$$yx^{-1} = a_i^{-1} g_i^{-1} h_i a_i \,, \quad i = 1, 2 \,. \tag{1.9.1}$$

This implies that yx^{-1} is a member of $U(a_1, a_2)$, the group of elements common to the groups $a_1^{-1} H_1 a_1$ and $a_2^{-1} H_2 a_2$. Given x, the h_i, and an element of $U(a_1, a_2)$, then the g_i and y are determined uniquely. Hence if there are $u(a_1, a_2)$ elements in $U(a_1, a_2)$ every pair T-similar to (a_1, a_2) will occur exactly $u(a_1, a_2)$ times in the above set. Thus there are

$$\frac{h^{(1)} h^{(2)} m!\, n!}{u(a_1, a_2)} \tag{1.9.2}$$

pairs in the equivalence class to which (a_1, a_2) belongs.

The pairs in the equivalence class with (a_1, a_2) contribute

$$\frac{h^{(1)} h^{(2)} m!\, n!}{u(a_1, a_2)} \cdot u(a_1, a_2) \tag{1.9.3}$$

to $\sum u(a_1, a_2)$, summed over the ordered pairs (a_1, a_2). As this is independent of the pair (a_1, a_2) it follows that

(1.9.4) $$\sum u(a_1, a_2) = h^{(1)} h^{(2)} m!\, n!\, (M \cdot N)_T \quad .$$

To evaluate $\sum u(a_1, a_2)$ we need to determine for each pair (a_1, a_2) the number of pairs of elements $v_1 \in H_1$ and $v_2 \in H_2$ such that

(1.9.5) $$a_1^{-1} v_1 a_1 = a_2^{-1} v_2 a_2 \quad .$$

Or we may choose the v_i first and then determine the number of pairs (a_1, a_2) satisfying (1.9.5). An alternate form of (1.9.5) is

(1.9.6) $$v_1 = a_1 a_2^{-1} v_2 a_2 a_1^{-1} = b^{-1} v_2 b \quad ,$$

where $b = a_2 a_1^{-1}$. For each element b satisfying (1.9.6) there are $m!\, n!$ pairs satisfying (1.9.5) ; namely $(a_1, b\, a_1)$ for each $a_1 \in S_m \times S_n$.

Any element of $S_m \times S_n$ may be interpreted as being a permutation of the elements of P performed in conjunction with a permutation of the elements of Q , where P and Q have been defined earlier. Since v_1 and v_2 are conjugate elements, by (1.9.6), it follows that they have the same cycle character with respect to the elements of P and Q , respectively. For fixed v_1 and v_2 , of type $(j) = (j_1, \ldots , j_m)$ with respect to P and of type $(\ell) = (\ell_1, \ldots , \ell_n)$ with respect to Q , the number of different solutions, b, to (1.9.6) is the number of ways of writing v_2 under v_1 , when both are expressed as a product of disjoint cycles, in such a way that all the cycles of

length one with respect to P come first, followed by the cycles of length two with respect to P, etc., and similarly for the cycles of permutations of elements of Q, or

$$(1.9.7) \qquad j_1! \ldots j_m!\, 2^{j_2} \ldots m^{j_m}\, \ell_1! \ldots \ell_n!\, 2^{\ell_2} \ldots n^{\ell_n} .$$

See e.g. Murnaghan [69], pp. 9-10.

Multiplying by $m!\, n!$ for reasons already given we see, since there are $h^{(1)}_{(j)(\ell)}\ h^{(2)}_{(j)(\ell)}$ pairs of elements $v_1 \in H_1$ and $v_2 \in H_2$ both of cycle type (j) and (ℓ) with respect to P and Q, respectively, using the notation of §1.8, that the total contribution to $\sum u(a_1, a_2)$ of the v_i's of type (j) and (ℓ) is

$$(1.9.8) \qquad m!\, n!\, h^{(1)}_{(j)(\ell)}\, h^{(2)}_{(j)(\ell)}\, T_{(j)}(m)\, T_{(\ell)}(n) ,$$

where $T_{(j)}(m) = j_1! \ldots j_m!\, 2^{j_2} \ldots m^{j_m}$ and similarly for $T_{(\ell)}(n)$. Summing this over the same partitions, (j) and (ℓ), as described for (1.3.4) and using (1.9.4) completes the proof of the following lemma.

Lemma 1.9.1. $\qquad (M \cdot N)_T = \dfrac{1}{h^{(1)}\, h^{(2)}} \sum_{(j)(\ell)} h^{(1)}_{(j)(\ell)}\, h^{(2)}_{(j)(\ell)}\, T_{(j)}(m)\, T_{(\ell)}(n) .$

When k-tuples of elements of $S_m \times S_n$ are being considered, $k \geq 2$, the T's are raised to the power $k-1$, besides the more obvious changes, since there would be $k-1$ equations corresponding to (1.9.6) all of which have the same number of solutions to be counted.

We may observe that the modified cycle indices of the groups H_1 and H_2, as defined in (1.8.1) contain all the information necessary to use this lemma.

We now develop the superposition theorem for bigraphs.

Let G_1 and G_2 be two m by n bigraphs each of whose point sets are labelled $P = \{P_1, \ldots, P_m\}$ and $Q = \{Q_1, \ldots, Q_n\}$. A superposition of G_1 and G_2 is defined as any bigraph which may be obtained by permuting the labels of the points of G_i according to any permutation

$$a_i = \begin{pmatrix} P_1 & P_2 & \cdots & P_m \\ P_{(1)a_i} & P_{(2)a_i} & \cdots & P_{(m)a_i} \end{pmatrix}\begin{pmatrix} Q_1 & Q_2 & \cdots & Q_n \\ Q_{(1)a_i} & Q_{(2)a_i} & \cdots & Q_{(n)a_i} \end{pmatrix},$$

where $a_i \in S_m \times S_n$, for $i = 1, 2,$ and then identifying all points in G_1 and G_2 which have the same label. In the superposed graph thus obtained we shall assume that the edges which come from G_1 are distinguishable from those which come from G_2, say from having a different colour.

Two different graphs both obtained by superpositioning G_1 and G_2 will be called <u>similar as labelled graphs</u> if, and only if, the same number of edges of both colours join P_i to Q_j in both graphs, $i = 1, \ldots, m$, and $j = 1, \ldots, n$; they will be called <u>isomorphic</u> if, and only if, the application of some permutation $a \in S_m \times S_n$ to the labels of the points of one of the graphs renders them similar as labelled graphs.

Those permutations of $S_m \times S_n$ which when applied to the labels of the points of G_i give a graph similar as a labelled graph to G_i form a group, the automorphism group, H_i, of G_i, for $i = 1, 2$. It follows that the two graphs obtained from G_i by applying the permutations a_i and b_i will be similar as labelled graphs if, and only if, $b_i a_i^{-1} \in H_i$, for $i = 1, 2$.

Any other graphs obtainable from G_1 and G_2 through an admissible permutation of the labels of the points may be indicated by specifying a pair of elements, (a_1, a_2), of $S_m \times S_n$. The superposed graphs corresponding to (a_1, a_2) and (b_1, b_2) will be similar as labelled graphs if, and only if, $b_i a_i^{-1} \in H_i$, for $i = 1, 2$, hence if, and only if, $(a_1, a_2) L (b_1, b_2)$.

Furthermore, the superposed graphs corresponding to (a_1, a_2) and (b_1, b_2) will be isomorphic if, and only if, there exists an element $x \in S_m \times S_n$ such that the graph corresponding to $(a_1 x, a_2 x)$ is similar as a labelled graph to the graph corresponding to (b_1, b_2), hence, if, and only if, $(a_1, a_2) T (b_1, b_2)$. But the number of T-inequivalent pairs has been determined already, which completes the proof of the superposition theorem for bigraphs.

Theorem 1.9.1. The number of nonisomorphic m by n bigraphs obtainable by superposing the m by n bigraphs G_1 and G_2 is the number $(M \cdot N)_T$, of Lemma 1.9.1, where H_1 and H_2 are the automorphism groups of G_1 and G_2 respectively.

In a similar fashion the more general theorem for superposing an arbitrary number of bigraphs comes from the more general form of the

lemma. Corresponding results could be obtained for superposing n-partite graphs, and the theorem is equally valid when the restriction that no more than one edge may join two points in different subsets is dropped.

1.10 The number of functional directed bigraphs

The counting series for the number of nonisomorphic functional digraphs, i.e. directed graphs in which each point is of outdegree one, was obtained by Harary [43] and later simplified by Read [79]. The method involved a repeated application of Pólya's theorem based upon an analysis of the structure of the digraph in terms of the length and number of the cycles it contained. (A cycle in a directed graph is the same as a cycle in an undirected graph except that its edges are all directed in the same sense and the possibility is also admitted of there being cycles of length two.) Essentially the same procedure, utilizing the results of §§ 1.4 and 1.8, would suffice to treat the corresponding problem for bigraphs. However, in this section we shall pursue an alternate approach making use of the superposition theorem for bigraphs. A more precise formulation of the problem follows.

Let there be given two distinct sets of points, $P = \{P_1, \ldots, P_m\}$ and $Q = \{Q_1, \ldots, Q_n\}$, $m,n \geq 1$, such that from each point in P there issues one, and only one, edge directed towards some point in Q, and similarly with P and Q interchanged. Such a configuration is called a <u>functional directed m by n bigraph</u> and may be interpreted as arising from an ordinary functional digraph in which the points are separated into two disjoint subsets and only points which belong to different subsets may be adjacent. The definition of isomorphism of

two such graphs follows that given in the preceding section as does the definition of the automorphism group of the graph.

It is natural to consider any functional directed bigraph as being a superposed graph in which the edges directed towards points in Q originate from one graph, call it G_1 , and the edges directed towards points in P originate from a second graph, call it G_2 . Hence, by Theorem 1.9.1, obtaining the modified cycle indices of $H^{(1)}$ and $H^{(2)}$, the automorphism groups of G_1 and G_2 , respectively, will enable us to determine $w_{m,n}$, the number of nonisomorphic m by n functional directed bigraphs.

The graph G_1 , containing m directed edges one from each P point towards a point of Q , may be characterized up to an isomorphism by a set of nonnegative integers, $(g) = (g_0, g_1, \ldots, g_m)$, where g_k is the number of points of Q of indegree k. It follows that

$$(1.10.1) \qquad g_0 + g_1 + \ldots + g_m = n$$

and

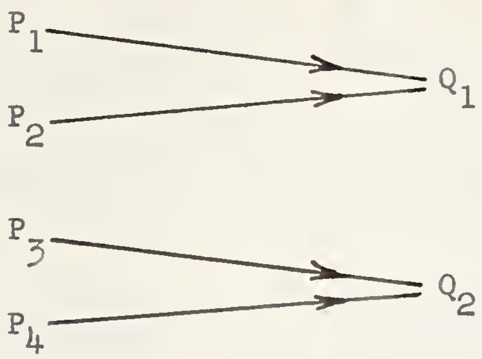
$$(1.10.2) \qquad 1 \cdot g_1 + 2 \cdot g_2 + \ldots + m \cdot g_m = m \, .$$

It is clear that under automorphisms of $H^{(1)}$ only those P points whose edge is directed towards the same Q point can be mapped into each other and that any point of Q can only be mapped into another point of Q whose indegree is the same. Hence, to determine $H^{(1)}$, or rather its modified cycle index, we may first consider the automorphism group of the subgraph consisting of the g_k Q points of indegree k , the $k \cdot g_k$ edges directed towards these points, and the

$k \cdot g_k$ P points from whence these edges are directed, for $0 \leq k \leq m$. The case $k = 0$ will be trivial.

Consider a g_k by k rectangular array in which the rows correspond to the g_k Q points of indegree k in some order and the entries in the rows represent the k P points, in some order, from which edges are directed towards the Q point to which the row corresponds. The group of the subgraph being considered then consists of any automorphism that may be obtained by first permuting the entries in each row according to some element of S_k, the symmetric group of degree k, not necessarily the same permutation for each row, and then permuting the rows themselves according to some element of Sg_k, the symmetric group of degree g_k. This group is known as the <u>composition</u>, $Sg_k[S_k]$, of Sg_k with S_k and the cycle index, with respect to the application of elements of the group to the P points involved, is obtained by "substituting" the cycle index of S_k into that of Sg_k, as shown in Pólya [74], p. 178. "Substituting" means that if $Z(Sg_k) = U(\varphi_1, \ldots, \varphi_{g_k})$ then φ_i is replaced by $V(\theta_i, \theta_{2i}, \ldots, \theta_{ki})$, where $V(\theta_1, \ldots, \theta_k) = Z(S_k)$. But the modified cycle index, $\bar{Z}(Sg_k[S_k])$, in which we are interested, indicates the cycle character of the elements of the automorphism group with respect to the corresponding points of Q as well. To accomplish this it is easily seen that we should replace each term in $Z(Sg_k)$ by the product of that term and the expression obtained from it by "substituting" in $Z(S_k)$. This enables us to obtain $\bar{Z}(Sg_k[S_k])$ from $Z(Sg_k)$ and $Z(S_k)$, both of which are known, in principle at least, by (1.2.5).

An example may clarify this. Let $g_2 = 2$. The configuration satisfying this has the following appearance:



Since $Z(S_2) = \frac{1}{2}(t_1^{\,2} + t_2)$ the modified cycle index of the automorphism group of this configuration is

$$\bar{Z}(S_2[S_2]) = \frac{1}{2}\left\{ [\tfrac{1}{2}(f_1^2 + f_2)]^2\, t_1^2 + [\tfrac{1}{2}(f_2^2 + f_4)]\, t_2 \right\}$$

$$= \frac{1}{8}(f_1^4\, t_1^2 + 2f_1^2\, f_2\, t_1^2 + f_2^2\, t_1^2 + 2f_2^2\, t_2 + 2f_4\, t_2)\ ,$$

as may be seen by inspection.

Applying the same argument for each value of k gives, according to Pólya [74], p. 177, the result that the modified cycle index of the automorphism group of G_1 corresponding to the set of numbers (g) is

$$(1.10.3) \qquad \bar{Z}\,(H^{(1)}_{(g)}) = \prod_{k=0}^{m} \bar{Z}\,(S_{g_k}\,[S_k])\ .$$

In the same way the graph G_2, containing n directed edges one from each point in Q to some point in P, may be characterized up to an isomorphism by a set of nonnegative integers, $(b) = (b_0, b_1, \ldots, b_n)$, where

$$(1.10.4) \qquad b_0 + b_1 + \ldots + b_n = m$$

and

$$(1.10.5) \qquad 1 \cdot b_1 + 2 \cdot b_2 + \ldots + n \cdot b_n = n \ .$$

We again find, with obvious notation, that

$$(1.10.6) \qquad \bar{Z}\,(H^{(2)}_{(b)}) = \prod_{\ell=0}^{n} \bar{Z}\,(S_{b_\ell}[S_\ell]) \ .$$

This would enable us to count the number of m by n functional directed bigraphs the indegrees of whose points satisfied the conditions implied by the values of the numbers in (g) and (b). To count all of them we should sum over all (g) and (b) satisfying (1.10.1) and (1.10.2), and (1.10.4) and (1.10.5), respectively. Using Theorem 1.9.1 this is summarized in the following theorem.

Theorem 1.10.1.

$$w_{m,n} = \sum_{(g),(b)} \left(\prod_{k=0}^{m} g_k!(k!)^{g_k} \prod_{i=0}^{n} b_i!(i!)^{b_i} \right)^{-1} \sum_{(j)(\ell)} h^{(g)}_{(j)(\ell)} \, h^{(b)}_{(j)(\ell)} \, T_{(j)}(m) T_{(\ell)}(n$$

where $h^{(g)}_{(j)(\ell)}$ and $h^{(b)}_{(j)(\ell)}$ are the number of automorphisms of type (j) with respect to the points of P and of type (ℓ) with respect to the points of Q in $H^{(1)}_{(g)}$ and $H^{(2)}_{(b)}$, respectively. The inner sum is over the same partitions as described for (1.3.4).

If $W(x,y) = \sum_{m,n=1}^{\infty} w_{m,n}\, x^m y^n$, then one finds using this theorem that

$$(1.10.7) \qquad \begin{aligned} W(x,y) = xy &+ xy^2 + xy^3 + xy^4 + xy^5 + \ldots \\ &+ 5x^2y^2 + 7x^2y^3 + 11x^2y^4 + 13x^2y^5 + \ldots \\ &+ 25x^3y^3 + 49x^3y^4 + 85x^3y^5 + \ldots \\ &+ 159x^4y^4 + 278x^4y^5 + \ldots \end{aligned}$$

CHAPTER II

ON BIPARTITE TOURNAMENTS

2.1 Definitions

A set of $n(\geq 1)$ points such that each pair of distinct points is joined by a single edge that is oriented towards exactly one of the points is called a (round-robin) tournament. One generalization of this definition is as follows: A set of $n(\geq 1)$ distinct subsets of points, $P_i = \{P_{i1}, \cdots, P_{in_i}\}$ with $n_i \geq 1$ for $i = 1, \cdots, n$, such that every pair of distinct points not belonging to the same subset is joined by an edge oriented towards exactly one of the points is called an n-partite tournament if there are no edges joining points in the same subset. If the edge joining P_{ij} to $P_{k\ell}$ is oriented towards the latter point this is indicated by writing $P_{ij} \to P_{k\ell}$ and similarly if the orientation is in the opposite sense. An ordinary tournament on n points is the same as an n-partite tournament with $n_1 = n_2 = \cdots = n_n = 1$.

In this chapter we investigate a few of the properties of n-partite tournaments and extend some of the results known about ordinary tournaments. The main emphasis is upon the case $n = 2$, bipartite tournaments. An m by n bipartite tournament may be more concisely defined as a complete oriented m by n bigraph, where a complete graph of any kind is one which contains as many edges as possible, consistent with its definition. Tournaments have been studied in connection with the method of paired comparisons (see e.g. Kendall [55] and David [12]) and a bipartite tournament provides a

model for the comparing of each member of one population with each member of a second population and making a decision, upon some predetermined basis, as to which of each pair is the preferred one.

To avoid repetition we remark at the outset that, unless otherwise stated, all statements of a probabilistic nature for bipartite tournaments are under the assumption that all possible outcomes in a given context are equally likely and that the orientations of the edges incident upon one point are independent of the orientations of the edges incident upon another point in the same subset.

2.2 Conditions for an oriented graph to be a bipartite tournament.

Consider an arbitrary oriented graph. We seek conditions under which the points of the graph may be assigned membership in one of two sets, P and Q, in such a way that the resulting configuration may be considered as being a bipartite tournament on the point sets P and Q. Clearly this property is independent of the orientation of the edges in the graph.

These conditions are contained in the following result.

Theorem 2.2.1. An arbitrary oriented graph, G, with V points and E edges may be considered as a bipartite tournament if, and only if, the graph G' obtained from G by ignoring the orientations of the edges satisfies the following two conditions:

(a) All cycles of G' are of even length.

(b) m points of G' are of degree $V - m$ and the remaining $V - m$ points are of degree m, where m is a nonnegative integer such that $m \leq V - m$. (If this is the case then $m = \frac{1}{2}(V - (V^2 - 4E)^{\frac{1}{2}})$.)

Any ordinary graph may be considered as being a bigraph if, and only if, its cycles are all of even length, by König [57], p. 151. The necessity of (b) as well is obvious.

To demonstrate the sufficiency of the conditions we observe first that (a) and König's theorem imply that the points of G' may be separated into two mutually exclusive and exhaustive subsets of $\overline{m}$ and $V - \overline{m}$ points, respectively, where $\overline{m} \leq V - \overline{m}$, such that no two points of the same subset are adjacent.

Assume $m > \overline{m}$. This and (b) imply that all the $V - m$ points of degree m must belong to the subset containing $\overline{m}$ points, since otherwise there wouldn't be enough points not in the same subset with them to enable them to have the specified degrees. But this implies that $V - m \leq \overline{m}$, or $V \leq m + \overline{m}$. It further implies that $m \leq V - \overline{m}$ or $m + \overline{m} \leq V$. Hence $\overline{m} = V - m \geq m$ and $m = V - \overline{m} \geq \overline{m}$, or $m = \overline{m}$, contradicting the assumption.

The assumption $m < \overline{m}$ yields the same contradiction, so we have as the only alternative that $m = \overline{m}$.

If $\overline{m} < V - \overline{m}$, then none of the points in the subset with $V - \overline{m}$ points can be of degree $V - m = V - \overline{m} > \overline{m}$. Hence all the points in the subset with $\overline{m}$ points must be of degree $V - \overline{m} = V - m$ and all the remaining points in the class of $V - \overline{m} = V - m$ points must be of degree $\overline{m} = m$.

If $\overline{m} = V - \overline{m} = V/2$ then every point is of degree $V/2$ and

since no point can be adjacent to any of the remaining $V/2 - 1$ points in the same subset as itself it must be adjacent to all the $V/2$ members of the other subset.

In either case restoring the original orientations to the edges yields a bipartite tournament by definition, completing the proof of the theorem.

To avoid the trivial case $\bar{m} = 0$ we need only demand that the original graph be nontrivial, i.e. that it contain at least one edge.

That there is no such simple characterization of those oriented graphs which may be considered as n-partite tournaments for $n > 2$ follows from the fact that there is no analogue to König's theorem for these cases, as cycles of any length (≥ 3) are possible in n-partite graphs for $n > 2$. (See Berge [3], p.32).

The total number of oriented graphs with $m(V - m)$ edges, $m \leq V - m$, and V points is

$$(2.1.1) \qquad \binom{\binom{V}{2}}{m(V - m)} \cdot 2^{m(V - m)} ,$$

the first factor being the number of ways of selecting the edges and the second being the number of ways of assigning the orientations. The number of these which may be regarded as being a bipartite tournament is

$$(2.1.2) \qquad \binom{V}{m} \cdot 2^{m(V - m)} ,$$

if $m < V - m$, and one half of this if $m = V - m$. Dividing (2.1.2) by (2.1.1) gives the probability that a random oriented graph with $m(V - m)$ edges and V points may be considered as being a bipartite

tournament.

Similarly one finds that the overall probability that a random oriented graph on V points may be considered as being a bipartite tournament is

$$3^{-\binom{V}{2}} \cdot \sum_{m=0}^{\frac{V-1}{2}} \cdot \binom{V}{m} \cdot 2^{m(V-m)} ,$$

if V is odd, and the sum should contain the additional term

$\binom{V}{V/2}\ 2^{(V/2)^2}\ \cdot\ \frac{1}{2}$ when V is even. The factor $3^{-\binom{V}{2}}$ arises from the fact that to determine any oriented graph on V points three choices must be made for each of the $\binom{V}{2}$ pairs of distinct points; which way to orient the edge joining the two points if they are to be joined or not to join them at all.

2.3 <u>On the score sequence of an n-partite tournament</u>.

By the <u>score</u> of P_{ij}, a point in an n-partite tournament is meant the outdegree of the point and will be denoted by v_{ij}. By the <u>soore sequence</u> of an n-partite tournament is meant the sets of scores $V_i = (v_{i1}, \cdots, v_{in_i})$, $i = 1, \cdots, n$. With no loss of generality we may assume that $v_{i1} \leq v_{i2} \leq \cdots \leq v_{in_i}$, for all i. In this section we give necessary and sufficient conditions for n sets of nonnegative integers to be the score sequence of an n-partite tournament.

These are contained in

Theorem 2.3.1. $n(\geq 1)$ sets of nonnegative integers

$V_i = (v_{i1}, \cdots, v_{in_i})$, when $v_{i1} \leq v_{i2} \leq \cdots \leq v_{in_i}$ and $n_i \geq 1$ for $i = 1, \cdots, n$, form the score sequence of some n-partite tournament if, and only if,

$$\text{(a)} \quad \sum_{i=1}^{n} \sum_{j=1}^{n_i} v_{ij} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} n_i n_j , \quad \text{and}$$

$$\text{(b)} \quad \sum_{i=1}^{n} \sum_{j=1}^{k_i} v_{ij} \geq \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_i k_j ,$$

for all sets of integers k_i, $i = 1, \cdots, n$, where $0 \leq k_i \leq n_i$, but not all $k_i = n_i$.

It is easily seen that (a) and (b) imply that

$$0 \leq v_{ij} \leq \left(\sum_{s=1}^{n} n_s \right) - n_i , \quad \text{for all } i \text{ and } j$$

involved.

That the conditions are necessary is obvious since (a) simply requires that the sum of all the scores be equal to the total number of edges in the configuration and (b) demands that the sum of the scores of the points in any proper subset of the tournament be at least as large as the number of oriented edges joining points in this subset.

Before proceeding to the proof of sufficiency it will be convenient to state a few preliminary results.

If the points of an n-partite tournament can be separated into two mutually exclusive and exhaustive subsets, A and B, such that every

edge which joins a point in A to a point in B is oriented towards the latter point then the configuration will be said to be reducible. To avoid making exceptions later the trivial 1-partite tournament will be considered reducible except when $n_1 = 1$. An n-partite tournament is irreducible if it is not reducible.

Dulmage and Mendelsohn [15] have applied the term irreducible to bigraphs with a different meaning. However no ambiguity should arise because the sense just described is the only one which will be used here. The concept of irreducibility of tournaments as defined here resembles somewhat the concept of irreducibility of a Markov chain, (See e.g. Feller [26], p. 349).

Lemma 2.3.1. An n-partite tournament is irreducible if, and only if, its score sequence satisfies condition (b) with strict inequality holding throughout.

This follows immediately from the definitions of the terms involved.

An n-partite tournament is defined as being strongly connected if, and only if, for each ordered pair of distinct points, P_{ij} and $P_{k\ell}$, there exists an oriented path from P_{ij} to $P_{k\ell}$, i.e. a sequence of distinct points whose first element is P_{ij}, whose last element is $P_{k\ell}$, and such that from each point in the sequence there issues an edge oriented towards its successor in the sequence.

The following statement is similar to a result published by Roy [86] but the argument is somewhat different.

Lemma 2.3.2. An n-partite tournament is strongly connected if, and only if, it is irreducible.

The necessity of the condition is obvious and if the tournament is not strongly connected there are two points, P_{ij} and $P_{k\ell}$, such that if B is the set of all points which can be reached from P_{ij} by some oriented path then $P_{k\ell}$ is in A, the set of all points of the tournament not in B. Moreover, any edge which joins a point in A to a point in B must be oriented towards the latter point, which implies reducibility by definition.

Straightforward considerations are sufficient to establish the following result.

Lemma 2.3.3. Given sets of integers, V_i, $i = 1, \cdots, n$, satisfying the conditions of Theorem 2.3.1, for which there exist integers, h_i, $0 \leq h_i \leq n_i$, for $i = 1, \cdots, n$, with not all $h_i = n_i$, such that equality holds in (b) when $k_i = h_i$. Then the sets $V_i' = (v_{i1}, \cdots, v_{ih_i})$ and the sets $V_i'' = (v_{ih_i+1} - b_i, \cdots\ v_{in_i} - b_i)$, where

$$b_i = \left(\sum_{s=1}^{n} h_s \right) - h_i ,$$

satisfy the conditions of Theorem 2.3.1 with n_i replaced by h_i and $n_i - h_i$, respectively for $i = 1, \cdots, n$.

Hence, if there exist two n-partite tournaments whose score sequences are the V_i''s and the V_i'''s, respectively, then combining them in a rather obvious fashion yields an n-partite tournament whose score sequence consists of the original sets, V_i, $i = 1, \cdots, n$. The interpretation is clear whenever any h_i equals 0 or n_i.

The proof of the sufficiency part of Theorem 2.3.1 now proceeds by multiple induction, first over the number of sets of points involved.

The case $n = 1$ is trivial, so we now assume that the theorem has been established for up to $n - 1$ sets of points, for some $n > 1$.

Suppose we have n sets, V_i, of integers satisfying the conditions of Theorem 2.3.1 in which $v_{n1} = v_{n2} = \cdots = v_{nn_n} = 0$. By the induction hypothesis and Lemma 2.3.3, with $h_1 = h_2 = \cdots = h_{n-1} = 0$ and $h_n = n_n$, we may assert the existence of an n-partite tournament whose score sequence consists of the V_i's.

For the second induction assume the required result has been proved for all cases where $0 \leq v_{nn_n} \leq t - 1$ and consider n sets, V_i, of integers which satisfy the conditions of Theorem 2.3.1 with $v_{nn_n} = t$ and with strict inequality holding throughout in (b).

Our third induction is now upon ℓ, the number of scores in V_n equal to t.

If $\ell = 1$ modify the given sets, V_i, by subtracting one from v_{nn_n}, which is greater than or equal to one by hypothesis, and by adding one to, say, the smallest v_{ij}, where $i < n$, in such a way as to leave all the numbers in each set still arranged in nondecreasing order. That is, if, say, $(v_{11}, v_{12}, v_{13}) = (0, 0, 1)$, one would be added to v_{12} and not to v_{11}. This change does not affect the total sum of the scores, so (a) is still satisfied. Also, any of the sums in (b) are decreased by at most one but as strict inequality held throughout

originally, by assumption, it follows that the modified sets of numbers still satisfy (b) as well.

The maximum score in V_n is now $t - 1$ so by the induction hypothesis for the second induction there exists an n-partite tournament whose score sequence is the set of numbers obtained by modifying the V_i's. We need now to change this n-partite tournament into one whose score sequence consists of the original V_i's.

Let P_{ij} be the point whose score was increased by one. Two cases now present themselves. If $P_{ij} \to P_{nn_n}$ in the tournament whose scores are the modified V_i's then simply reversing the orientation of the edge joining P_{ij} and P_{nn_n} yields a configuration whose score sequence is the original set of V_i's. If, however, $P_{nn_n} \to P_{ij}$ another approach is required.

If the n-partite tournament assured us by the induction hypothesis is reducible, then P_{nn_n} must be in the set called B in the definition of reducibility. This follows from Lemma 2.3.1 and the fact that any partial sum of the numbers in the modified set of V_i's giving equality in (b) must have included the score of P_{nn_n}. P_{ij} must also be in B since $P_{nn_n} \to P_{ij}$. But this is impossible, using Lemma 2.3.1 again, since any partial sum in (b) of the numbers in the modified set of V_i's which includes the scores of both P_{ij} and P_{nn_n} is unchanged from what it was before any of the numbers were changed; hence strict inequality would hold and the only alternative, under our assumptions, is that if $P_{nn_n} \to P_{ij}$ then the tournament is irreducible.

By Lemma 2.3.2 there exists an oriented path from P_{ij} to P_{nn_n}. The only effect upon the scores caused by reversing the sense of orientation of all edges in this path will be to decrease P_{ij}'s by one and to increase P_{nn_n}'s by one.

In either case we have produced an n-partite tournament whose score sequence is the original set of V_i's.

We now drop the assumption that strict inequality held throughout in (b). What has already been shown along with the application of Lemma 2.3.3, repeated if necessary, completes the proof for the case $\ell = 1$.

Now assume that the assertion has been demonstrated for $v_{nn_n} = t$ and up to $\ell = r - 1 \geq 1$. We consider sets of numbers V_i, $i = 1, \cdots, n$, in which the last r elements of V_n are equal to t, and which satisfy the conditions of Theorem 2.3.1 with strict inequality holding throughout in (b). The same scheme only this time reducing $v_{nn_n-(r-1)}$ by one suffices to complete the proof of the third induction.

But this completes also the proof of the induction on the maximum score in the n th set of numbers, which shows that the theorem is valid for n sets of numbers, which completes the argument for the original assertion, by induction.

This theorem is actually only a special case of a more general result on simultaneous solutions in integers to set-theoretic functions satisfying certain conditions.

Corollary 2.3.1. A set of integers $u_1 \leq u_2 \leq \cdots \leq u_n$ is the score

sequence of some ordinary round-robin tournament if, and only if,

$$\sum_{i=1}^{k} u_i \geq \binom{k}{2} ,$$

for $k = 1, 2, \cdots, n$, with equality holding when $k = n$.

This was first proved by Landau [61], without using the concept of irreducibility, and follows immediately from Theorem 2.3.1 upon setting $n_1 = n_2 = \cdots = n_n = 1$.

Corollary 2.3.2. There exists an m by n matrix of 0's and 1's with row sums r_i, where $r_1 \leq r_2 \leq \cdots \leq r_m$, and column sums c_j , where $c_1 \geq c_2 \geq \cdots \geq c_n$, if, and only if

$$\sum_{i=1}^{k} r_i + \sum_{j=1}^{\ell} (m - c_j) \geq k\ell ,$$

for $k = 0, 1, \cdots, m$ and $\ell = 0, \cdots, n$, with equality holding when $k = m$ and $\ell = n$.

The proof of this involves applying Theorem 2.3.1 to assert the equivalence of the hypothesis and the existence of an m by n bipartite tournament with score sequence $V_1 = (r_1, \cdots, r_m)$ and $V_2 = (m-c_1, \cdots, m-c_n)$ and constructing the required matrix, $A = \| a_{ij} \|$, by letting $a_{ij} = 1$ or 0 according as $P_{1i} \to P_{2j}$ or $P_{2j} \to P_{1i}$, respectively.

We may observe that in checking, for fixed k , if

$$\sum_{i=1}^{k} r_i + \sum_{j=1}^{\ell} (m - c_j) \geq k\ell ,$$

it is necessary to consider only those ℓ for which $m - c_\ell < k$, for if it is satisfied for these values it will automatically hold for the remaining cases. A similar remark may be made for fixed values of ℓ.

Given two sets of nonnegative integers, r_i, $i = 1, \cdots, m$, and c_j, $j = 1, \cdots, n$, with $r_1 \leq r_2 \leq \cdots \leq r_m \leq n$ and $m \geq c_1 \geq c_2 \geq \cdots \geq c_n$. Let S_j denote the j th column sum of an m by n matrix whose i th row consists of r_i 1's followed by $n - r_i$ 0's. A theorem discovered independently by Gale [28] and Ryser [87] states that the r_i and c_j can be the row and column sums, respectively, of an m by n matrix of 0's and 1's if, and only if,

$$c_1 + c_2 + \cdots + c_j \leq S_1 + S_2 + \cdots + S_j ,$$

for $j = 1, \cdots, n$, with equality holding for $j = n$.

Assume that a given set of r_i and c_j satisfy the conditions of Corollary 2.3.2. For any fixed value of ℓ, $\ell = 1, \cdots, n$, let k be the number such that $r_k \leq \ell$ and $r_{k+1} > \ell$. If no such k exists take $k = m$.

Then

$$(r_1 + r_2 + \cdots + r_k) + ((m-c_1) + (m-c_2) + \cdots + (m-c_\ell)) \geq k\ell ,$$

for any suitable value of ℓ, by Corollary 2.3.2. Or,

$$\begin{aligned}(c_1 + c_2 + \cdots + c_\ell) &\leq (r_1 + r_2 + \cdots + r_k) + \ell(m - k) \\ &= S_1 + S_2 + \cdots + S_\ell ,\end{aligned}$$

appealing to the definition of S_j and k, with equality holding for $\ell = n$.

Making use of the remark following Corollary 2.3.2 the converse implication may be demonstrated in a similar fashion. Hence the Gale-Ryser Theorem and Corollary 2.3.2 are equivalent.

2.4 <u>The number of bipartite tournaments</u>.

The determination of the number of nonisomorphic m by n bipartite tournaments, $B_{m,n}$, may be considered as a special case of the counting problems previously considered. It was described in § 1.3 how to obtain the counting series for the number of nonisomorphic oriented m by n bigraphs. The coefficient of the last term, x^{mn}, is clearly the number of nonisomorphic m by n bipartite tournaments.

An alternate approach is to observe that there is a one-to-one correspondence between m by n bipartite tournaments with t edges oriented towards points in Q and $mn - t$ edges oriented towards points in P and ordinary m by n bigraphs with t edges. Hence, the total number of nonisomorphic m by n bipartite tournaments is the sum of the coefficients in $g_{m,n}(x)$ or $g_{m,n}(1)$. Rather than evaluate this counting series and then substitute in $x = 1$ it is simpler to let $x = 1$ when first substituting the figure counting series into the cycle index. This gives the result that

$$B_{m,n} = Z(G_{m,n}; 2) = \frac{1}{m!n!} \sum_{(j),(\ell)} h^{(m)}_{(j)}\, h^{(n)}_{(\ell)}\, 2^{\Sigma j_a \ell_b (a,b)} , \tag{2.4.1}$$

where the summation is over the same partitions and indices as described for (1.3.4).

The first few terms of $B(x,y) = \sum_{m,n=0}^{\infty} B_{m,n}\, x^m y^n$ are found to

be as follows:

$$
\begin{aligned}
B(x,y) = x &+ 2xy + 3xy^2 + 4xy^3 + 5xy^4 + 6xy^5 + \cdots \\
&+ 7x^2y^2 + 13x^2y^3 + 22x^2y^4 + 34x^2y^5 + \cdots \\
&+ 36x^3y^3 + 87x^3y^4 + 190x^3y^5 + \cdots \\
&+ 317x^4y^4 + 1053x^4y^5 + \cdots \\
&+ 5554x^5y^5 + \cdots \quad .
\end{aligned} \tag{2.4.2}
$$

The form of Pólya's theorem used in (2.4.1) which counts the total number of inequivalent configurations regardless of their content was found independently by Davis [13] and Slepian [91]. For a further discussion of the relationship between this form and Pólya's original version see Harary [39].

The largest term in (2.4.1), corresponding to $(j) = (m)$ and $(\ell) = (n)$, gives the inequality

$$B_{m,n} \geq 2^{mn}/m!n! \tag{2.4.3}$$

It seems likely that for large m and n and with some restriction upon their relative orders the term retained to give this inequality represents an increasingly larger proportion of all m by n bipartite tournaments. A better inequality, resulting from considering also the partitions $(j) = (m-2,1)$ and $(\ell) = (n-2,1)$ for $m, n \geq 2$, is

$$B_{m,n} \geq \frac{1}{m!n!} \left(2^{mn} + \binom{m}{2} 2^{(m-1)n} + \binom{n}{2} 2^{(n-1)m} + \binom{m}{2}\binom{n}{2} 2^{(m-1)(n-1)}\right) \tag{2.4.4}$$

If the remaining terms are to become negligible in comparison with the first term then certainly the logarithm of m must become negligible

in comparison with n, and vice versa, as m and n become large. For $m = n = 5$ these two bounds are approximately .42 and .76, respectively, of $B_{5.5}$.

For similar remarks with respect to ordinary tournaments see Landau [59] and Davis [14]. For more general approximations and references to other relevant papers see Harary [41].

It should be observed that determining the number of nonisomorphic m by n bipartite tournaments is not equivalent to determining the number of different score sequences for such tournaments, for while having the same score sequence is obviously a necessary condition for two bipartite tournaments to be isomorphic it is not sufficient as may be seen by the example to be given in § 2.12. The same applies to ordinary tournaments, as pointed out by Landau [59]. The determination of the number of score sequences appears to be unsolved, although Mohanty and Narayana [64] claim to have some upper bounds for the number of score sequences of ordinary tournaments and David [12] contains tables and frequencies of the score sequences of tournaments involving fewer than nine points, among other things.

An obvious, albeit virtually useless, scheme for generating all the score sequences of m by n bipartite tournaments is by the exponents of the quantities in the terms of the expansion of the following product:

$$\prod_{\substack{i=1,\cdots,m \\ j=1,\cdots,n}} (P_i + Q_j) \quad .$$

2.5 The number of irreducible bipartite tournaments.

By determining all the values of k and ℓ for which equality holds in the sum of Corollary 2.3.2 it is easily seen how every m by n bipartite tournament may be formed from irreducible m_v by n_v bipartite tournaments, $U_{m_v n_v}$, where $v = 1, 2, \cdots, h$, h a positive integer, $m_1 + m_2 + \cdots + m_h = m$, and $n_1 + n_2 + \cdots + n_h = n$, by letting each P point and Q point in $U_{m_v n_v}$ be joined by an edge to each Q point and P point, respectively, in $U_{m_{v+1} n_{v+1}}, \cdots, U_{m_h n_h}$, for $v = 1, 2, \cdots, h - 1$, which is oriented towards the latter point. Furthermore, starting with a given m by n bipartite tournament the graphs $U_{m_v n_v}$, $v = 1, \cdots, h$, called the irreducible bipartite sub-tournaments of the original one, are uniquely determined up to an isomorphism. (If the original graph is irreducible then $h = 1$ and $U_{m_1 n_1}$ is the graph itself). The converse statement is valid also. This observation enables us to count the number of nonisomorphic irreducible bipartite tournaments in terms of the total number of non-isomorphic bipartite tournaments.

Let $b_{m,n}$ and $A_{m,n}$ denote the number of nonisomorphic m by n bipartite tournaments which are irreducible and reducible, respectively, and let their corresponding generating functions be $b(x,y)$ and $A(x,y)$. We recall that $b_{0,1} = b_{1,0} = 1$ and $b_{0,i} = b_{i,0} = 0$, for $i = 2, 3, \cdots$, from the definition of irreducibility given in § 2.3. Then

$$(2.5.1) \qquad b(x,y) + A(x,y) = B(x,y) ,$$

where $B(x,y)$ is as defined in the preceeding section, since every bipartite tournament is either reducible or irreducible, but not both. Also

$$(2.52) \qquad B(x,y) = b(x,y) + b^2(x,y) + b^3(x,y) + \cdots = \frac{b(x,y)}{1 - b(x,y)} \quad ,$$

since the k th power of $b(x,y)$, for $k = 1, 2, \cdots,$ enumerates those bipartite tournaments which may be considered as having been formed by combining k irreducible bipartite tournaments according to the scheme described above.

Solving for $b(x,y)$ gives

$$(2.5.3) \qquad b(x,y) = \frac{B(x,y)}{1 + B(x,y)} \quad ,$$

which could have been derived directly by use of the principle of inclusion and exclusion.

From (2.51) and (2.53) we have

$$(2.5.4) \qquad A(x,y) = \frac{B^2(x,y)}{1 + B(x,y)} = b(x,y)\, B(x,y) \quad ,$$

or equivalently

$$(2.5.5) \qquad A(x,y) = \frac{b^2(x,y)}{1 - b(x,y)} \quad .$$

It is easily seen that these could have been derived in other orders.

By restricting the number of terms that appear in $b(x,y)$ it is not difficult to count bipartite tournaments in terms of other parameters, e.g. in terms of the size of the largest irreducible subtournament they

contain.

Similar considerations suffice to count the number of nonisomorphic irreducible ordinary tournaments in terms of the total number of ordinary tournaments, which may be regarded as known, from Davis [14] and Harary [40]. Counting ordinary tournaments in terms of the largest irreducible subtournament contained is the same as counting them in terms of the longest oriented cycle they contain since for ordinary tournaments irreducibility is equivalent to the existence of an oriented cycle through all the points from results of Camion [6] and Roy [86].

In a paper on weighted compositions of integers Moser and Whitney [68] have considered certain other generating functions related by equations similar to (2.5.2) and (2.5.3).

2.6 The probability that a random bipartite tournament is irreducible.

It has been shown elsewhere [65] that if $Q(n)$ denotes the probability that an ordinary round-robin tournament on n points chosen at random from the set of $2^{\binom{n}{2}}$ such tournaments is reducible, then

$$(2.6.1) \qquad \left| Q(n) - \frac{2n}{2^{n-1}} \right| < \frac{1}{2^{n-1}}, \quad \text{for} \quad n \geq 8.$$

In this section we extend the methods used to obtain this to derive information about $P(m,n)$, the probability that an m by n bipartite tournament chosen at random from the set of 2^{mn} such tournament is irreducible. $P(1,0) = P(0,1) = 1$ and $P(i,0) = P(0,i) = 0$, for $i = 2, 3, \cdots,$ by definition.

If a given bipartite tournament is reducible then it contains a

maximal, in a sense, proper subset of points such that these points together with the edges joining them form an irreducible bipartite subtournament and all edges joining points not in this subset with points which are in it are oriented towards the latter points. This is a consequence of the considerations in the first paragraph of the preceeding section where this subgraph was denoted by $U_{m_h n_h}$.

The probability that this subgraph contains k P points and ℓ Q points, $1 \leq k \leq m$, $1 \leq \ell \leq n$, but not both $k = m$ and $\ell = n$, is

$$\binom{m}{k} \binom{n}{\ell} P(k,\ell)\ 2^{k\ell}\ 2^{(m-k)(n-\ell)} / 2^{mn}$$

$$= \frac{\binom{m}{k} \binom{n}{\ell} P(k,\ell)}{2^{k(n-\ell)}\ 2^{\ell(m-k)}},$$

since $P(k,\ell)\ 2^{k\ell}$ irreducible k by ℓ tournaments may be formed on each of the $\binom{m}{k} \binom{n}{\ell}$ subsets of k P points and ℓ Q points, and having chosen one of these there remain only $(m-k)(n-\ell)$ edges whose sense of orientation needs to be chosen.

The only other alternative for reducible m by n tournaments, $m, n \geq 1$, is that there exists a subset of points, of necessity all belonging to the same set, P or Q, whose scores are zero. In this case $U_{m_h n_h}$ would not necessarily be uniquely determined, although it would be determined up to an isomorphism. The probability of this being the case is

$$\left[1 - (1 - 1/2^n)^m \right] + \left[1 - (1 - 1/2^m)^n \right].$$

Combining these cases and summing over k and ℓ in the range described above gives the result that

$$(2.6.2) \qquad P(m,n) = 1 - \left[1 - (1 - 1/2^n)^m \right] - \left[1 - (1 - 1/2^m)^n \right]$$

$$- \sum \frac{\binom{m}{k} \binom{n}{\ell} \; P(k,\ell)}{2^{k(n-\ell)} \; 2^{\ell(m-k)}} \quad .$$

Since $P(k,\ell) = 0$ if k or ℓ equals one we need sum only over $k = 2, \ldots, m$ and $\ell = 2, \cdots, n$, but not both $k = m$ and $\ell = n$.

Using this the first few nontrivial values of $P(m,n)$ are found to be approximately as follows:

(2.6.3)

$P(2,2) = .12500$,

$P(3,2) = .09375$, $P(3,3) = .19921$,

$P(4,2) = .05462$, $P(4,3) = .22140$, $P(4,4) = .23624$,

$P(5,2) = .02921$, $P(5,3) = .19867$, $P(5,4) = .43096$, $P(5,5) = .76173$.

As the computational labor involved in using (2.6.2) very soon becomes prohibitive we seek some estimates for the value of $P(m,n)$. We make the further restriction that $m = n$.

The following inequality is easily established.

Lemma 2.6.1. $\quad m^3 \leq (m - 3)\, 2^{2(m-1)}$, for integers $m \geq 4$.

Lemma 2.6.2. $\quad m^3 \leq (m - 3)\,(m - t)\, 2^{m-1+t}$, for integers $m \geq 4$.
and $t = 2, 3, \ldots, m - 2$.

When $m = 4$ it may be verified directly. We next show that the inequality is valid for $t = 2$ and all admissible values of m.

Assume that

$$k^3 \leq (k-3)(k-2)\,2^{k+1}, \quad \text{for some integer } k \geq 4 .$$

Using the induction hypothesis and obvious inequalities gives

$$\begin{aligned}
(k+1)^3 &= k^3 + 3k^2 + 3k + 1 \\
&\leq (k-3)(k-2)2^{k+1} + \left[\frac{3}{2}\left(\frac{k-3}{k}\right)(k-2)2^{k+1}\right. \\
&\quad \left. + \frac{3}{2}(k-3)\left(\frac{k-2}{k}\right)2^{k+1}\right] + \left[3\left(\frac{k-3}{k}\right)\left(\frac{k-2}{k}\right)2^{k+1}\right] + 1 \\
&\leq (k-3)(k-2)2^{k+1} + (k-2)2^{k+2} + (k-3)2^{k+2} + \\
&\qquad (k-3)(k-2)2^{k+1} + 2^{k+2} \\
&= \left[(k+1) - 3\right]\left[(k+1) - 2\right] 2^{k+2}
\end{aligned}$$

The proof of the lemma will be completed if we can now show that the right hand side is an increasing function of t, for $2 \leq t \leq m-2$.

This is equivalent to showing that the derivative with respect to t of the right hand side is positive in this range, or that

$$(m-3)\left[-2^{m-1+t} + (m-t)\,2^{m-1+t}\log 2\right] > 0 .$$

This will be so if

$$m - t > \frac{1}{\log 2},$$

or

$$t < m - \frac{1}{\log 2} \quad ,$$

which is certainly the case from the restrictions on t. This completes the proof of the lemma.

We now extend this bound for m^3 to include another variable.

Lemma 2.6.3 $m^3 \leq (m-3)(m-j)(m-\ell)\, 2^{j+\ell}$, for integers $m \geq 4$, and $j, \ell = 2, \ldots, m-1$.

The inequality is easily verified when $m = 4$ and all admissible j and ℓ. Assume that it has been demonstrated for all values of m up to $k \geq 4$, and all admissible values of j and ℓ for each such values of m. Then for fixed j and ℓ between 2 and $k-1$ it follows that

$$\begin{aligned}
(k+1)^3 &= k^3 + 3k^2 + 3k + 1 \\
&\leq (k-3)(k-j)(k-\ell)2^{j+\ell} + \Big[\left(\tfrac{k-3}{k}\right)(k-j)(k-\ell) \\
&\quad + (k-3)\left(\tfrac{k-j}{k}\right)(k-\ell) + (k-3)(k-j)\left(\tfrac{k-\ell}{k}\right) \Big] 2^{j+\ell} \\
&\quad + \Big[(k-3)\left(\tfrac{k-j}{k}\right)\left(\tfrac{k-\ell}{k}\right) + \left(\tfrac{k-3}{k}\right)(k-j)\left(\tfrac{k-\ell}{k}\right) + \left(\tfrac{k-3}{k}\right)\left(\tfrac{k-j}{k}\right)(k-\ell) \Big] 2^{j+\ell} + 1 \\
&\leq \Big[(k-3)(k-j)(k-\ell) + (k-j)(k-\ell) + (k-3)(k-\ell) \\
&\quad + (k-3)(k-j) + (k-3) + (k-j) + (k-\ell) + 1 \Big] 2^{j+\ell} \\
&= \big[(k+1) - 3 \big] \big[(k+1) - j \big] \big[(k+1) - \ell \big] 2^{j+\ell}
\end{aligned}$$

The only values of j and ℓ, in the case $m = k+1$, not included in this inductive argument are when one of j or ℓ but not

both equals k, in which case the result follows from Lemma 2.6.2, or when $j = \ell = k$, in which case the result follows from Lemma 2.6.1. This completes the proof of the lemma by induction.

Lemma 2.6.4 $$\frac{\binom{m}{j}\binom{m}{\ell}}{2^{j(m-\ell)}2^{\ell(m-j)}} \leq \frac{1}{m(m-1)(m-2)},$$

for integers $m \geq 16$, and $j, \ell = 2, \ldots, m$, but not $j = \ell = m$.

Straightforward, but tedious, calculations suffice to show that the inequality is valid for $m = 16$ and all admissible j and ℓ.

Assume that it has been verified for all values of m up to $k \geq 16$ and all admissible values of j and ℓ for each value of m. Let j and ℓ be fixed between 2 and k but not with both equalling k. Then

$$\frac{\binom{k+1}{j}\binom{k+1}{\ell}}{2^{j(k+1-\ell)}2^{\ell(k+1-j)}} = \frac{(k+1)^2}{(k+1-j)(k+1-\ell)2^{j+\ell}} \cdot \frac{\binom{k}{j}\binom{k}{\ell}}{2^{j(k-\ell)}2^{\ell(k-j)}}$$

$$\leq \frac{(k+1)^2}{(k+1-j)(k+1-\ell)2^{j+\ell}} \cdot \frac{1}{k(k-1)(k-2)} \leq \frac{(k+1)^2}{(k+1)^3\, k(k-1)}$$

$$= \frac{1}{(k+1)[(k+1)-1][(k+1)-2]},$$

using the induction hypothesis and Lemma 2.6.3.

The remaining values that j and ℓ may assume, when $m = k+1$, for which the validity of the lemma is not established in this argument are as follows:

(i) $j = \ell = k$.

Verifying the inequality in this case is equivalent to showing that $(k+1)^3(k)(k-1) \leq 2^{2k}$, for $k \geq 16$, which is easily seen to be valid for $k \geq 8$.

(ii) $j = k + 1$ and $\ell = 2, \ldots, k$.

To show the correctness of the lemma when $j = k + 1$ and $\ell = k$ is equivalent to showing that $(k+1)^2(k)(k-1) \leq 2^{k+1}$, for $k \geq 16$, which also may be done in a straightforward manner.

Since

$$\frac{\binom{k+1}{\ell+1}}{2^{(k+1)[(k+1)-(\ell+1)]}} = \frac{2^{k+1}(k+1-\ell)}{\ell + 1} \cdot \frac{\binom{k+1}{\ell}}{2^{(k+1)(k+1-\ell)}}, \quad \text{for } \ell = 2,\ldots,k-1,$$

and

$$\frac{2^{k+1}(k+1-\ell)}{\ell + 1} > 1, \quad \text{for } k \geq 16,$$

it follows that in this case the left member of the inequality is an increasing function of ℓ for fixed k. Hence its maximum over the range under consideration occurs at $\ell = k$ where the validity of the inequality has already been established.

Appealing to symmetry to treat the remaining cases completes the proof of the lemma by induction.

Since $P(k,\ell) \leq 1$ and the sum in (2.6.2) contains $m(m-2)$ terms Lemma 2.6.4 may be used to give the following result:

Theorem 2.6.1 $\quad 2(1-1/2^m)^m - 1 - \frac{1}{m-1} < P(m,m) < 2(1-1/2^m)^m - 1,$

for $m \geq 16$.

This can be used to derive an asymptotic estimate for $P(m,m)$ and, in particular, shows that $P(m,m)$ tends to one as m tends to infinity.

In the next section a partial answer will be given to the question of for what relative orders of m and n is it still true that $P(m,n)$ tends to one as m and n tend to infinity. That this is not always the case is easily seen. For an extreme example let $m \geq \log\left(\frac{1}{\epsilon}\right) 2^n$, where $0 < \epsilon < 1$, as m and n tend to infinity. Then

$$\begin{aligned} P(m,n) &< (1 - 1/2^n)^m + (1 - 1/2^m)^n - 1 \\ &< (1 - 1/2^n)^m < e^{-\frac{m}{2^n}} \leq e^{\log \epsilon} = \epsilon \ . \end{aligned}$$

Moser [67] has shown that if the maximum score of any point in an ordinary tournament on n points is less than or equal to $3/4(n-1)$ and if the minimum score is greater than or equal to $1/4(n-1)$ then the tournament is irreducible.

For bipartite tournaments there is the following similar result:

Theorem 2.6.2. If the scores of the P and Q points of an m by n bipartite tournament all lie strictly between $n/4$ and $3n/4$, and $m/4$ and $3m/4$, respectively then the tournament is irreducible.

The proof will be by contradiction. If the bipartite tournament is reducible then its points may be separated into two mutually exclusive and exhaustive classes, A and B, such that every edge which joins a point in A to a point in B is oriented towards the latter point. Let B contain k P points and ℓ Q points, and A contain $m - k$ P points and $n - \ell$ Q points where k and ℓ, considering nontrivial cases only, are integers such that $1 \leq k \leq m - 1$ and $1 \leq \ell \leq n - 1$.

If $k \leq m/4$, $k \geq 3m/4$, $\ell \leq n/4$, or $\ell \geq 3n/4$ then it is easily seen that some score violates the hypothesis of the theorem. If both $k \leq m/2$ and $\ell \leq n/2$ then the minimum score is certainly less than or equal to $k/2 \leq m/4$, if it belongs to a Q point, or $\ell/2 \leq n/4$, if it belongs to a P point, contradicting the hypothesis of the theorem in either case. A similar contradiction results from considering the maximum score when both $k \geq m/2$ and $\ell \geq n/2$.

By symmetry there remains to consider only the case where, say, $m/2 \leq k < 3m/4$ and $n/4 < \ell \leq n/2$, but not both $k = m/2$ and $\ell = n/2$.

Assuming that the scores satisfy the hypothesis of the theorem it follows that the total number of edges joining two points in A is strictly less than

$$(3m/4 - k)(n - \ell) \quad + \quad (3n/4 - \ell)(m - k) \; .$$

If this quantity can be shown to be less than or equal to $(m - k)(n - \ell)$, the actual number of edges joining two points in A , a contradiction will have been derived.

Similarly, still assuming that the scores satisfy the hypothesis of the theorem, the total number of edges joining two points in B is strictly greater than

$$kn/4 \; + \; \ell m/4 \; .$$

If this quantity can be shown to be greater than or equal to $k\ell$, the actual number of edges joining two points in B , another contradiction will have been derived.

The inequalities which we would like to verify under the conditions on k and ℓ become, upon simplifying and rearranging:

$$(2.6.4) \qquad n(m/2 - 3k/4) \leq \ell(3m/4 - k) \quad ,$$

and

$$(2.6.5) \qquad \ell\, m/4 \geq k\,(\ell - n/4) \quad .$$

Replacing ℓ by $n/4$ in (2.6.4) the resulting inequality is seen to be valid when $k \geq 5m/8$; hence if $n/4 < \ell \leq n/2$ and $5m/8 \leq k < 3m/4$ (2.6.4) holds, implying a contradiction. We need now consider only the case where $m/2 \leq k < 5/8\, m$.

Replacing k by $5/8m$ in (2.6.5) the resulting inequality is seen to be valid when $\ell \leq 5n/12$; hence if $n/4 < \ell \leq 5n/12$ and $m/2 \leq k < 5m/8$ (2.6.5) holds, implying a contradiction. We need now consider only the case where $5n/12 < \ell \leq n/2$.

The new lower bound for ℓ is now substituted into (2.6.4) again and the resulting inequality is valid for $k \geq 9m/16$, so henceforth we may assume that $m/2 \leq k < 9m/16$. Using this bound for k in (2.6.5) we find that the resulting inequality is now valid for $\ell \leq 9n/20$, so henceforth we may assume that $9n/20 < \ell \leq n/2$.

By induction it follows that after the t th iteration of this process (2.6.4) will have been established for $n/4 < \ell \leq n/2$ and $\frac{4t+1}{8t} \cdot m \leq k < 3m/4$ and (2.6.5) will have been established for $m/2 \leq k < 3m/4$ and $n/4 < \ell \leq \frac{4t+1}{8t+4} \cdot n$, which will suffice to prove the theorem for these cases.

For fixed m and n only a finite number of such iterations are

necessary to bring the above bounds on k and ℓ, for which the theorem is valid, to within, say, a distance of one-half from $m/2$ and $n/2$, respectively. Having already disposed of the case where $k = m/2$ and $\ell = n/2$ this suffices to complete the proof of the original assertion since k and ℓ are integers.

2.7 On the diameter of a bipartite tournament.

If a and b are two distinct points in an oriented graph then by the distance from a to b is meant the length of the shortest oriented path from a to b, where the length of a path is the number of edges it contains. If no such path exists we shall say that the distance is infinite. The distance from a to b is not necessarily the same as the distance from b to a under this definition. By the diameter of an oriented graph is meant the maximum value of the distance between any two distinct points of the graph. From Lemma 2.3.2 it follows that the diameter of a reducible n-partite tournament is infinite.

For some results on the diameter of ordinary directed graphs see Ghouila-Houri [29].

Moser [67] has shown that the probability that an ordinary tournament on n points has diameter two tends to one as n tends to infinity; furthermore, if $k = o(n/\log n)$, as k and n tend to infinity, then a random tournament on n points will have more than k paths of length two joining every pair of distinct points with probability tending to one. Nevertheless, there exist tournaments with finite diameters as large as $n - 1$, for $n = 3, 4, \ldots$. In this section some related results are derived for bipartite tournaments.

Since some of the statements of the type we would like to prove

for m by n bipartite tournaments are not necessarily true when there is an extreme disparity in the relative magnitudes of m and n we shall make the restriction for the remainder of this section that m and n tend to infinity in such a way that

$$(2.7.1) \qquad \log m = o(n) \quad \text{and} \quad \log n = o(m) \quad .$$

In the first place it is easily seen that the probability of there being any point of indegree zero in a random m by n bipartite tournament is less than or equal to $m/2^n + n/2^m$, which tends to zero as m and n tend to infinity while satisfying (2.7.1).

In the second place the probability that in a random m by n tournament there exists two distinct points in P, P_i and P_j, such that there is no path of length two from P_i to P_j, i.e. no point Q_ℓ such that $P_i \to Q_\ell$ and $Q_\ell \to P_j$, is at most

$$\frac{m(m-1)\, 3^n\, 2^{mn-2n}}{2^{mn}} = m(m-1)\,(3/4)^n \quad ,$$

since there are $m(m-1)$ ways of choosing the ordered pair of P points, three admissible ways of orienting the two edges joining these two points to each of the n Q points, and two ways of orienting each of the remaining $mn-2n$ edges. But this quantity also tends to zero as m and n tend to infinity while satisfying (2.7.1). This and a corresponding argument with the two P points replaced by two Q points implies that the distance betwen any two distinct points both of which are in the same point set of a bipartite tournament is two, with probability tending to one.

Since it is impossible for a bipartite tournament to have diameter equal to one or two, the observations in the two preceeding paragraphs imply the following theorem.

Theorem 2.7.1. The probability that an m by n bipartite tournament has diameter three tends to one as m and n tend to infinity while satisfying (2.7.1).

Corollary 2.7.1. The probability that an m by n bipartite tournament is irreducible tends to one as m and n tend to infinity while satisfying (2.7.1).

This provides a partial answer to the question mentioned in the prece~~e~~ding section following Theorem 3.6.1, and is an immediate consequence of Theorem 2.7.1 and the earlier remark that the diameter of a reducible tournament is infinite.

The probability that there are not more than k paths of length two joining every ordered pair of P points, P_i and P_j, is at most

$$\frac{m(m-1)}{2^{mn}} \binom{n}{k} 4^k 3^{n-k} 2^{m n-2 n} = m(m-1) \binom{n}{k} (3/4)^{n-k} ,$$

by a slight extension of the previous considerations. It is not difficult to see that this tends to zero as m and n tend to infinity while satisfying (2.7.1) if $k = o(n/\log n)$. A corresponding result applies when the two P points are replaced by two Q points, providing a property of bipartite tournaments similar to that of ordinary tournaments mentioned above.

Deeper considerations would presumably enable one to make stronger statements concerning the number of paths of length three which join two points not in the same point set of a bipartite tournament.

2.8 A hierarchy index for bipartite tournaments.

In studying the structure of animal societies in which a non-transitive dominance relation exists between every pair of distinct members, such as the peck-right in a flock of hens, Landau [59], [60] has considered what he called a hierarchy index which provides a measure of how closely the society resembles a hierarchy with respect to the dominance relation. A hierarchy with respect to a given dominance relation is a society whose members can be so ordered that each member dominates all those who are beneath him in the ordering and is dominated by all those who are above him in the ordering. An ordinary tournament provides a model of such a structure and in this section we consider an analogue of Landau's hierarchy index for bipartite tournaments although a natural interpretation for the situation here is perhaps not as readily available as that for ordinary tournaments.

In a nontrivial m by n bipartite tournament in which the scores of the points P_i and Q_j are denoted by v_i and u_j, respectively, $i = 1, \ldots, m$, $j = 1, \ldots, n$, the hierarchy index, h, is defined by

$$h = \frac{4}{m\,n\,(m+n)} \left[\sum_{i=1}^{m} (v_i - n/2)^2 + \sum_{j=1}^{n} (u_j - m/2)^2 \right]. \tag{2.8.1}$$

This index ranges from zero, when all $v_i = n/2$ and all $u_j = m/2$ which can happen only if both m and n are even, and one, when either

all of the edges are oriented towards points of Q or they are all oriented towards points of P. This situation corresponds to a hierarchy in the ordinary case. The factor $\frac{4}{m\,n\,(m+n)}$ is chosen so as to make the index have the value one in the latter case.

We shall determine the mean and variance of h under the assumptions given in § 2.1 .

Denoting the mean and variance of a random variable, y, by $E(y)$ and $\sigma^2(y)$, respectively, it follows that

$$E(h) = \frac{4}{m\,n\,(m+n)} \left\{ \sum_{i=1}^{m} E\left[(v_i - n/2)^2\right] + \sum_{j=1}^{n} E\left[(u_j - m/2)^2\right] \right\}$$

$$= \frac{4}{m\,n\,(m+n)} \left\{ m\, E\left[(v_i - n/2)^2\right] + n\, E\left[(u_i - m/2)^2\right] \right\},$$

since the expectation of a sum equals the sum of the expectations of the terms in the sum and the quantities in the same sum in the above expression are equal. But $E\left[(v_1 - n/2)^2\right]$ is simply the variance of the sum of n independent random variables which have the same binomial distribution with mean $1/2$, and is therefore $n/4$. Similarly $E\left[(u_1 - m/2)^2\right] = m/4$. Substituting these into the above equation gives

$$(2.8.2) \qquad E(h) = 2/(m+n) .$$

The determination of the variance is somewhat more involved and may proceed as follows:

$$\sigma^2(h) = E(h^2) - E^2(h)$$

$$= \left[\frac{4}{mn(m+n)} \right]^2 E \left[\left(\sum_{i=1}^{m} (v_i - n/2)^2 + \sum_{j=1}^{n} (u_j - m/2)^2 \right)^2 \right] - E^2(h)$$

$$= \left[\frac{4}{mn(m+n)} \right]^2 \Big\{ m\, E[(v_1-n/2)^4] + m(m-1)\, E[(v_1-n/2)^2(v_2-n/2)^2]$$

$$+ 2mn\, E[(v_1-n/2)^2\, (u_1-m/2)^2]$$

$$+ n(n-1)\, E[(u_1-m/2)^2(u_2-m/2)^2] + n\, E[(u_1-m/2)^4] \Big\}$$

$$- E^2(h) \,,$$

upon expanding the first expression, taking expectations, and combining those terms which are equal.

$E[(v_1-n/2)^4] \;=\; n(3n-2)/16.$ (See e.g. Kendall and Stuart [56], p. 121.)

$E[(v_1-n/2)^2(v_2-n/2)^2] = E^2[(v_1-n/2)^2] = n^2/16$, since the orientations of the edges incident upon P_1 are independent of the orientations of the edges incident upon P_2.

Finally, $E[(v_1-n/2)^2\, (u_1-m/2)^2] = E[(v_1-n/2)^2] \cdot E[(u_1-m/2)^2] = mn/16.$ One way of establishing this is to replace v_1 and u_1 by $\bar{v}_1 + \delta_1$ and $\bar{u}_1 + \delta_2$, respectively, where δ_1 is one or zero according as to whether $P_1 \rightarrow Q_1$ or $Q_1 \rightarrow P_1$, respectively, and $\delta_2 = 1 - \delta_1$. $\bar{v}_1$ represents the score of P_1 with respect to the points $Q_2, \ldots, Q_n$ and a corresponding remark applies to $\bar{u}_1$. The statement follows upon expanding the resulting expression, observing that $\bar{v}_1$ and $\bar{u}_1$ are independent of each other, that $E(\delta_1) = E(\delta_2) = 1/2$ and that $E(\delta_1\, \delta_2) = 0$, and making use of known values of the moments of the binomial distribution.

Substituting these values and those obtained by symmetry into the expression for $\sigma^2(h)$ and simplifying one has the following result:

$$(2.8.3) \qquad \sigma^2(h) = \frac{2(m+n-2)}{(m+n)^2\, mn} \; .$$

By Chebyshev's inequality it follows that as m and n tend to infinity the probability that h differs very much from zero tends to zero.

More general situations in which the probabilities that $P_i \to Q_j$ and $Q_j \to P_i$ are denoted by p_{ij} and q_{ji}, respectively, where $0 \leq p_{ij} \leq 1$ and $p_{ij} + q_{ji} = 1$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, may be considered. In this case the mean and variance of h may be calculated, in principle at least, by using characteristic functions.

Other quantities could be used to define a hierarchy index but that in (2.8.1) appears to be as simple as any. For example, the $m/2$ and $n/2$ appearing in the sums of (2.8.1) could both be replaced by $mn/(m+n)$ and the factor in front modified so that the resulting expression still varies between zero and one. However the mean and variance now become quite complicated functions of m and n and in particular don't necessarily tend to zero as m and n tend to infinity. Further results about the distribution of h will arise from a somewhat different context in the next section.

We conclude this section with a simple example of the distribution of h when $m = 3$ and $n = 2$ under the hypothesis of randomness mentioned earlier. Since h is a function of the scores the various

score sequences, the probability that a random 3 by 2 bipartite tournament has such a score sequence, and the corresponding values of h are listed, from which the expected value and variance of h may be calculated for this case.

Distribution of h for $m = 3$ and $n = 2$

$V = (v_1, v_2, v_3)$; $U = (u_1, u_2)$	$2^6 \cdot$Probability of V and U	15 h
(2,2,2); (0,0)	1	15
(1,2,2); (0,1)	6	9
(0,2,2); (1,1)	3	7
(1,1,2); (1,1)	6	3
(1,1,2); (0,2)	6	7
(0,1,2); (1,2)	12	5
(1,1,1); (0,3)	2	9
(1,1,1); (1,2)	6	1
(0,1,1); (1,3)	6	7
(0,1,1); (2,2)	6	3
(0,0,2); (2,2)	3	7
(0,0,1); (2,3)	6	9
(0,0,0); (3,3)	1	15

$E(h) = 2/5$

$\sigma^2(h) = 1/25$

2.9 On a measure associated with a bipartite tournament

In this section we consider another quantity which provides, in a sense, a measure of how nearly alike the various points of a bipartite tournament are with respect to the orientations of the edges incident upon them. This quantity will admit of a perhaps more intuitive interpretation than did the hierarchy index but we will find that the two are very closely related.

It will be convenient to introduce first the definition of the adjacency matrix of a bipartite tournament, essentially given already in the proof of Corollary 2.3.2.

Given an m by n tournament whose point sets are $P = \{P_1, P_2, \ldots, P_m\}$ and $Q = \{Q_1, \ldots, Q_n\}$; by the adjacency matrix of this tournament is meant the m by n matrix, $A = \|a_{ij}\|$, where a_{ij} is one or zero according as to whether $P_i \to Q_j$ or $Q_j \to P_i$, respectively. Later we shall need the n by m matrix, $C = \|c_{ji}\|$, where c_{ji} is one or zero according as to whether $Q_j \to P_i$ or $P_i \to Q_j$, respectively. Clearly, A plus the transpose of C equals the m by n matrix all of whose entries are ones.

The binary distance, $b_{k\ell}$, between the distinct points P_k and P_ℓ of a bipartite tournament is defined as follows:

$$(2.9.1) \qquad b_{k\ell} = \sum_{j=1}^{n} |a_{kj} - a_{\ell j}| ,$$

where $A = \|a_{ij}\|$ is the adjacency matrix of the tournament as described above. $b_{k\ell}$ is simply the number of points Q_i such that the edges joining P_k and P_ℓ to Q_i have different orientations. The term "binary distance" is used because the definition of $b_{k\ell}$ is essentially the same as the definition of the distance between two binary words in information theory. (See e.g. Peterson [73], p. 7.)

g_{st} , the binary distance between the distinct points Q_s and Q_t , is similarly defined by

$$(2.9.2) \qquad g_{st} = \sum_{i=1}^{m} |a_{is} - a_{it}| .$$

In an m by n tournament let b denote the minimum binary distance between any two distinct P points.

The number of m by n bipartite tournaments having $b = 0$, or equivalently, the number of m by n matrices of 1's and 0's having at least two rows identical, is at most $\binom{m}{2} 2^n 2^{n(m-2)}$, since for each of the $\binom{m}{2}$ choices of two distinct rows there are 2^n ways of causing them to be identical and there remain $n(m-2)$ entries to be chosen either zero or one. The ratio of this number to 2^{mn}, the total number of m by n tournaments is less than $m^2/2^{n+1}$ which tends to zero provided that m and n satisfy (2.7.1) as they tend to infinity.

More generally, the number of m by n bipartite tournaments having $b \leq k$ is less than or equal to the number of m by n adjacency matrices having two rows whose entries differ in at most k places which is at most $\binom{m}{2} 2^n \binom{n}{k} 2^k 2^{n(m-2)}$, since having chosen the two rows in one of $\binom{m}{2}$ ways and filling out the entries of one of them in 2^n ways we may then choose in one of $\binom{n}{k}$ ways the k columns in which the other row may possibly differ, the remaining n-k entries being determined by the corresponding entries in the row already filled out. The entries in these k positions in the second row may be chosen in one of 2^k ways and there still remain $n(m-2)$ entries which may be zero or one. The ratio of this number to the total number of m by n tournaments is less than $m^2 n^k/2^{n+1-k}$. As m and n tend to infinity while satisfying (2.7.1) it is not difficult to see that this quotient tends to zero if $k = o(n/\log n)$. These observations suffice to prove the following result:

Theorem 2.9.1. If $k = o(n/\log n)$ the probability that the minimum binary distance between two distinct P points of an m by n bipartite tournament is less than or equal to k tends to zero as m and n tend to infinity while satisfying (2.7.1).

By symmetry similar results hold with respect to the minimum binary distance between two distinct Q points. Also, the same type of argument may be used to derive corresponding results on the maximum binary distance between two points.

A quantity, D, which provides a measure of the degree of similarity subsisting between the points of an m by n bipartite tournament may be defined as follows:

$$(2.9.3) \qquad D = \sum_{k,\ell} b_{k\ell} + \sum_{s,t} g_{st} ,$$

where the sums are over all unordered pairs of integers between 1 and m and 1 and n, respectively. An immediate consequence of the definitions of the terms involved is that $D = 0$ if, and only if, all the entries of the incidence matrix associated with the bipartite tournament are the same.

As before let v_i and u_j denote the scores of points P_i and Q_j, respectively, in an m by n tournament, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Then the ith row sum and the jth column sum of the incidence matrix of such a tournament are v_i and $m-u_j$, respectively. Hence, the ith row, containing v_i 1's and $n-v_i$ 0's, contributes $v_i(n-v_i)$ to Σg_{st} and similarly the jth column contributes $u_j(m-u_j)$ to $\Sigma b_{k\ell}$. Summing over the rows and columns gives an alternative expression for D.

$$(2.9.4) \qquad D = \sum_{i=1}^{m} v_i(n-v_i) + \sum_{j=1}^{n} u_j(m-u_j)$$

$$= n \sum_{i=1}^{m} v_i + m \sum_{j=1}^{n} u_j - \sum_{i=1}^{m} v_i^2 - \sum_{j=1}^{n} u_j^2 \quad .$$

In this form it is easily seen that the maximum value D may assume occurs when the v_i and u_j are all as nearly equal to $n/2$ and $m/2$, respectively, as is possible. Straightforward considerations for the various cases imply the following results:

Theorem 2.9.2.

$$D \leq \begin{cases} \frac{m\,n}{4}(m+n) \quad , & \text{if } m \equiv n \equiv 0(2) \text{ ;} \\ \frac{mn-1}{4}(m+n) \quad , & \text{if } m \equiv n \equiv 1(2) \text{ ;} \\ \frac{m\,n}{4}(m+n) - \frac{m}{4} \quad , & \text{if } m \equiv 0(2) \text{ and } n \equiv 1(2) \text{ ;} \\ \frac{m\,n}{4}(m+n) - \frac{n}{4} \quad , & \text{if } m \equiv 1(2) \text{ and } n \equiv 0(2) \text{ .} \end{cases}$$

Rather obvious examples can be constructed for which equality holds for each possibility.

Expanding the right member of

$$\frac{m\,n}{4}(m+n)\,h = \sum_{i=1}^{m} (v_i-n/2)^2 + \sum_{j=1}^{n} (u_j-m/2)^2$$

and comparing it with (2.9.4) shows that

$$(2.9.5) \qquad \frac{m\,n}{4}(m+n)\,h + D = \frac{m\,n}{4}(m+n) \text{ .}$$

In other words the hierarchy index, as defined in (2.8.1), apart from a constant factor for fixed values of m and n, and D, as defined in (2.9.3), measure the same thing except for a difference in origin, as it were. Hence, the results of §2.8 can be adapted to apply to D and, conversely, what we shall presently show about the distribution of D can be modified to give corresponding results about the distribution of h in perhaps an easier manner than it would have been to have derived them for h directly.

Since the expected values of each of the terms in the sums in (2.9.3) are $n/2$ and $m/2$, respectively, it follows that

$$(2.9.6) \qquad E(D) = \frac{m\,n}{4}(m+n-2) ,$$

under the hypothesis of randomness.

The variance of D may be calculated by methods similar to those which will be used to calculate the third moment about the mean or it may be obtained almost directly from (2.8.3) and (2.9.5). Either approach gives

$$(2.9.7) \qquad \sigma^2(D) = \frac{m\,n}{8}(m+n-2) .$$

In obtaining an approximation to the distribution of D under the given assumptions we shall require the third moment, $\mu_3(D)$, of D about its mean. First we obtain an expression for the expected value of D^3.

Raising both sides of (2.9.3) to the third power, taking expectations, and combining terms of the same type gives

$$E(D^3) = \binom{m}{2} E(b_{12}^3) + 3 \binom{m}{2}\binom{m-2}{2} E(b_{12}^2\, b_{34}) \tag{2.9.8}$$

$$+ 3m(m-1)(m-2)\, E(b_{12}^2\, b_{13}) + 3 \binom{m}{2}\binom{n}{2} E(b_{12}^2\, g_{12})$$

$$+ \binom{m}{2}\binom{m-2}{2}\binom{m-4}{2} E(b_{12}\, b_{34}\, b_{56}) + 3\binom{m}{2}\binom{m-2}{2}\binom{n}{2} E(b_{12}\, b_{34}\, g_{12})$$

$$+ 6m \binom{m-1}{2}\binom{m-3}{2} E(b_{12}\, b_{13}\, b_{45}) + 6m \binom{m-1}{3} E(b_{12}\, b_{13}\, b_{14})$$

$$+ 6\binom{m}{2}(m-2)(m-3)\, E(b_{12}\, b_{13}\, b_{34}) + 6m \binom{m-1}{2}\binom{n}{2} E(b_{12}\, b_{13}\, g_{12})$$

$$+ 6\binom{m}{3} E(b_{12}\, b_{23}\, b_{13}) + (\text{terms obtained by interchanging}$$

m and n and b_{ij} and g_{ij}).

The expected values of the various terms are calculated separately using the independence of various variables whenever possible.

(i) $E(b_{12}^3) = \frac{n^2}{8}(n+3)$, from Kendall and Stuart [56], p. 58. (There appears to be a misprint in the second equation from the bottom of the page in the last reference. It reads

$$\mu_3' = n(n-1)(n-2)\, p^3 + 3n(n-1)\, p^2 - 2np ,$$

but from the process of derivation it is easily seen that the last -2 should be a $+1$ which gives the result needed here.)

(ii) $E(b_{12}^2\, b_{34}) = E(b_{12}^2)\, E(b_{34}) = (\frac{n}{4} + \frac{n^2}{4})(\frac{n}{2}) = \frac{n^2}{8}(n+1)$,

using properties of the binomial distribution.

(iii) $E(b_{12}^2\, b_{13}) = E(b_{12}^2)\, E(b_{13}) = \frac{n^2}{8}(n+1)$, since if the first row of the adjacency matrix is thought of as having been filled in the

expression becomes a function of the entries in the second and third rows which are independent of each other.

a_{11}, a_{12}, a_{21}, and a_{22}, in the adjacency matrix of a bipartite tournament, may be all the same. If this is not the case it may be that entries in the same row are alike, that entries in the same column are alike, that diagonally opposite entries are alike, or that three of the entries are of one kind and the fourth is of the other kind. The corresponding probabilities of these events are found to be 1/8, 1/8, 1/8, 1/8, and 1/2, respectively. Calculating the expected value of $b_{12}^2\ g_{12}$ for the various cases gives the result that

$$\text{(iv)}\quad E(b_{12}^2\ g_{12}) = \frac{1}{8}\left[\frac{m-2}{4} + \left(\frac{m-2}{2}\right)^2\right]\left(\frac{n-2}{2}\right)$$

$$+\frac{1}{8}\left\{\left[\frac{m-2}{4} + \left(\frac{m-2}{2}\right)^2\right] + 4\left(\frac{m-2}{2}\right) + 4\right\}\left(\frac{n-2}{2}\right)$$

$$+\frac{1}{8}\left[\frac{m-2}{4} + \left(\frac{m-2}{2}\right)^2\right]\left(\frac{n-2}{2} + 2\right) + \frac{1}{8}\left\{\left[\frac{m-2}{4} + \left(\frac{m-2}{4}\right)^2\right]\right.$$

$$\left. + 4\left(\frac{m-2}{4}\right) + 4\right\}\left(\frac{n-2}{2} + 2\right)$$

$$+\frac{1}{2}\left\{\left[\frac{m-2}{4} + \left(\frac{m-2}{2}\right)^2\right] + 2\left(\frac{m-2}{2}\right) + 1\right\}\left(\frac{n-2}{2} + 1\right) = \frac{mn}{8}(n+1)\ .$$

That is, $E(b_{12}^2\ g_{12}) = E(b_{12}^2)\ E(g_{12})$, which may not be intuitively obvious.

$$\text{(v)}\quad E(b_{12}\ b_{34}\ b_{56}) = E(b_{12})\ E(b_{34})\ E(b_{56}) = \frac{n^3}{8}\ .$$

$$\text{(vi)}\quad E(b_{12}\ b_{34}\ g_{12}) = E(b_{12})\ E(b_{34})\ E(g_{12}) = \frac{n^2 m}{8}\ ,$$ from either the type of procedure used in (iv) or that to be illustrated for (xi).

(vii) $E(b_{12}\, b_{13}\, b_{45}) = E(b_{12})\, E(b_{13})\, E(b_{45}) = \frac{n^3}{8}$.

(viii) $E(b_{12}\, b_{13}\, b_{14}) = E(b_{12})\, E(b_{13})\, E(b_{14}) = \frac{n^3}{8}$.

(ix) $E(b_{12}\, b_{13}\, b_{34}) = E(b_{12})\, E(b_{13})\, E(b_{34}) = \frac{n^3}{8}$.

(x) $E(b_{12}\, b_{13}\, g_{12}) = E(b_{12})\, E(b_{13})\, E(g_{12}) = \frac{n^2 m}{8}$.

Statements (vii) - (x) follow by reasoning similar to that employed in establishing (iii) and (vi) .

To obtain the expected value of the remaining term we observe that we may write

$$b_{12} = t_{12}^{(1)} + t_{12}^{(2)} + \dots + t_{12}^{(n)} \text{ , where}$$

$$t_{12}^{(j)} = |a_{1j} - a_{2j}| \text{ , for } j = 1, \dots, n \text{ ,}$$

and similarly for the other two factors. Upon substituting these expressions into the product, $b_{12}\, b_{23}\, b_{13}$, expanding, taking expectations, and combining similar terms, we have that

$$\begin{aligned} E(b_{12}b_{23}b_{13}) &= n(n-1)(n-2)\, E(t_{12}^{(1)} t_{23}^{(2)} t_{13}^{(3)}) \\ &+ 3n(n-1)\, E(t_{12}^{(1)} t_{23}^{(1)} t_{13}^{(2)}) \\ &+ n\, E(t_{12}^{(1)} t_{23}^{(1)} t_{13}^{(1)}) \text{ .} \end{aligned}$$

But $E(t_{12}^{(1)} t_{23}^{(2)} t_{13}^{(3)}) = E(t_{12}^{(1)})\, E(t_{23}^{(2)})\, E(t_{13}^{(3)}) = \frac{1}{8}$, since these are independent, involving, as they do, different columns. Also, $E(t_{12}^{(1)} t_{23}^{(1)} t_{13}^{(2)}) = E(t_{12}^{(1)} t_{23}^{(1)})\, E(t_{13}^{(2)})$, by independence. Considering

the various combinations of values which occur in the first product shows that the expected value of the complete product is $1/4 \cdot 1/2$. The consideration of these same possibilities shows that $E(t_{12}^{(1)}\ t_{23}^{(1)}\ t_{13}^{(1)}) = 0$. Substituting these expected values gives

$$\text{(xi)} \qquad E(b_{12}\ b_{23}\ b_{13}) = \frac{n}{8}\,(n^2-1).$$

Substituting the values in (i) - (xi) plus those obtained by symmetry into (2.9.8) gives an expression for $E(D^3)$.

But $\mu_3(D) = E(D^3) - 3E(D)\,\sigma^2(D) - E^3(D)$. Substituting (2.9.6), (2.9.7), and (2.9.8) into this equation and simplifying the expression thus obtained gives the following result:

$$\text{(2.9.9)} \qquad \mu_3(D) = \frac{-mn}{8}\,[(m-1)(m-2) + (n-1)(n-2)]\ .$$

From equations (2.9.5) and the one preceding we see that D is related to something having the appearance of a variable which would have the χ^2- distribution were all the terms involved statistically independent, which they aren't. Nevertheless we can transform D into a variable whose first three moments are those of a χ^2- distribution.

The first three moments of the χ^2- distribution,

$$d\,F \propto e^{-\frac{1}{2}\chi^2}\ \chi^{\nu-1}\ d\chi\ ,$$

are given by $\mu_1' = \nu$, $\mu_2 = 2\nu$, and $\mu_3 = 8\nu$. (See e.g. Kendall and Stuart [56], p. 370).

$$\text{(2.9.10)} \qquad x = k\,[\frac{mn}{4}\,(m+n-2) - D] + \nu$$

has mean ν also.

Since $\mu_2(Ax + B) = A^2 \mu_2(x)$ for fixed A and B we must have

$$k = (2\nu)^{\frac{1}{2}} \left(\frac{mn}{8}(m+n-2)\right)^{-\frac{1}{2}}$$

in order for the variance of x to be equal to 2ν. We must also have

$$\frac{\frac{mn}{8}[(m-1)(m-2) + (n-1)(n-2)]}{[\frac{mn}{8}(m+n-2)]^{3/2}} = \frac{8\nu}{(2\nu)^{3/2}}$$

or

$$\nu = \frac{mn(m+n-2)^3}{[(m-1)(m-2) + (n-1)(n-2)]^2} ,$$

since $\mu_3(Ax + B) = A^3 \mu_3(x)$ for fixed A and B.

Substituting into (2.9.10) implies that

$$(2.9.11) \qquad x = \frac{4(m+n-2)}{(m-1)(m-2) + (n-1)(n-2)} \left[\frac{mn}{4}(m+n-2) - D\right] + \nu$$

has its first three moments the same as a χ^2- distribution with ν degrees of freedom.

As an example we consider the case $m = n = 4$. By direct enumeration it may be seen that, barring errors, 3042 of the 2^{16} 4 by 4 bipartite tournaments give a value of D less than or equal to 16. Denoting the probability of the occurrence of the event A by $Pr(A)$ we have, approximately, that $Pr(D \leq 16) = 0.046$ in this case.

(2.9.11) becomes $x = 2[24 - D] + 24$ and from tables of the χ^2- distribution with 24 degrees of freedom we find, approximately, that $Pr(x \geq 40) = 0.022$.

It is not difficult to see that $D \neq 17$ for any 4 by 4 tournament. Hence, the accuracy of the approximation for this case may be improved by replacing D by $D+1$, where the 1 represents the so-called correction for continuity. (For a discussion of this see Kendall [55], p. 42 and p. 135, where the same procedure as used here is carried out for a different problem for ordinary tournaments.) Doing this the probability that D is less than or equal to 16 should be near to the probability that x, as defined above, is greater than or equal to 38. From tables this latter probability is found to be approximately 0.038 which is somewhat nearer to the actual value, 0.046.

Presumably the agreement would be better for larger values of m and n.

2.10 <u>The number of acyclic bipartite tournaments</u>

A tournament which contains no oriented cycles is said to be <u>acyclic</u>. Up to an isomorphism there is only one ordinary tournament on n points which is acyclic. This is the graph called a hierarchy in §2.8. If the points of the tournament are labelled then clearly there are $n!$ ways of ranking them so as to form a hierarchy. The counting of the number of acyclic bipartite tournaments is not quite as trivial and will be carried out in this section. Both the labelled and unlabelled cases will be treated.

If in a bipartite tournament all of the points had an outdegree greater than zero one could go indefinitely from point to point along an oriented path. Since the graph contains only a finite number of points this path would eventually pass through the same point twice,

thus forming a cycle. Hence, every acyclic bipartite tournament contains some points of outdegree zero which, of necessity, must all be in the same point set, say, P. With this assumption a moment's reflection suffices to establish the validity of the following assertion, since the removal of points of outdegree zero and the edges incident upon them from an acyclic bipartite tournament leaves an acyclic graph remaining. In every acyclic m by n tournament, in which the points of outdegree zero are points of P, the points of P and Q may be split into k and ℓ mutually exclusive and exhaustive non-vacuous classes, where the ith and jth class contain r_i and s_j points, respectively,

$$(2.10.1) \qquad k-\ell = 1 \text{ or } 0 , \quad \sum_{i=1}^{k} r_i = m , \quad \sum_{j=1}^{\ell} s_j = n ,$$

and such that the edges joining the points in the ith class of P points to the points in the first $i-1$ classes of Q points, $i = 2, \ldots, k$, are oriented towards the points in Q and all the remaining edges are oriented towards points in P. Furthermore, each such acyclic bipartite tournament determines uniquely the r_i's and s_j's, in order, and the converse as well is true up to an isomorphism. Therefore, the number of nonisomorphic acyclic m by n tournaments in which the points of outdegree zero are points of P is the number of ordered sets of strictly positive integers satisfying (2.10.1). The remaining case may be treated by symmetry.

If $k = \ell = r$, $r = 1, 2, \ldots, \min[m,n]$, the number of solutions to (2.10.1) is the product of the number of compositions

of m and n, respectively, into r parts, or (see Riordan [84], p. 124)

$$\binom{m-1}{r-1}\binom{n-1}{r-1}.$$

If $k-1 = \ell = r$, $r = 1, 2, \ldots, \min[m-1,n]$, the number of solutions to (2.10.1) is the product of the number of compositions into $r+1$ and r parts of m and n, respectively, or

$$\binom{m-1}{r}\binom{n-1}{r-1}.$$

Upon interchanging m and n to treat the case where the points of outdegree zero are Q points and summing over r we have the following result, where $K_{m,n}$ denotes the number of nonisomorphic acyclic m by n tournaments.

Theorem 2.10.1.

$$K_{m,n} = 2\sum_{r=1}^{\min[m,n]}\binom{m-1}{r-1}\binom{n-1}{r-1} + \sum_{r=1}^{\min[m-1,n]}\binom{m-1}{r}\binom{n-1}{r-1}$$
$$+ \sum_{r=1}^{\min[m,n-1]}\binom{m-1}{r-1}\binom{n-1}{r}.$$

Corollary 2.10.1.

$$[\pi(m-1)]^{-\frac{1}{2}}\, 2^{2m-1} \sim 2\binom{2(m-1)}{m-1} \le K_{m,m} \le 2\binom{2m}{m} \sim (\pi m)^{-\frac{1}{2}}\, 2^{2m+1}.$$

Setting $m = n$ in Theorem 2.10.1 gives

$$K_{m,m} = 2 \left\{ \sum_{r=1}^{m-1} \binom{m-1}{r-1} \left[\binom{m-1}{r-1} + \binom{m-1}{r} \right] + \binom{m-1}{m-1}^2 \right\}$$

$$= 2 \sum_{r=1}^{m} \binom{m-1}{r-1} \binom{m}{r} \leq 2 \sum_{r=0}^{m} \binom{m}{r}^2 = 2\binom{2m}{m}$$

$$\sim (\pi\, m)^{-\frac{1}{2}} \; 2^{2m+1} \quad ,$$

using well known binomial identities and Stirling's formula. (See e.g. Feller [26], pp. 62-63.) The other half of the corollary may be shown in a similar fashion.

Next we determine $L_{m,n}$, the number of acyclic m by n tournaments with labelled points. The number of such tournaments, in which the points of outdegree zero are P points, and in which $k = \ell = r$, $r = 1, 2, \ldots, \min[m,n]$, where k and ℓ have the same meaning as before, is the product of the number of ways of putting m and n different things into r different cells, respectively, with no cell empty, or (see Riordan [84], p. 91)

$$r!\, S(m,r)\; r!\, S(n,r) \; ,$$

where $S(m,j)$ denotes a Stirling number of the second kind defined by

$$t^m = \sum_{j=0}^{m} S(m,j)\, t_{(j)} \; , \quad m > 0 \; .$$

Analogous results for the other cases are obtained in a similar fashion and combining them gives

Theorem 2.10.2.

$$L_{m,n} = 2 \sum_{r=1}^{\min[m,n]} (r!)^2 \, S(m,r) \, S(n,r)$$

$$+ \sum_{r=1}^{\min[m-1,n]} (r+1)! \; r! \; S(m,r+1) \, S(n,r)$$

$$+ \sum_{r=1}^{\min[m,n-1]} (r+1)! \; r! \; S(m,r) \, S(n,r+1) \; .$$

Corollary 2.10.2.

$$L_{m,m} = 2 \sum_{r=1}^{m} (r!)^2 \, S(m+1,\, r+1) \, S(m,r) \; .$$

The proof of this is similar to part of that of Corollary 2.10.1 except that the identity (see Riordan [84], p. 33) used here is

$$S(m,r) + (r+1) \, S(m,r+1) = S(m+1,\, r+1) \; .$$

2.11 <u>A condition on the score sequence of a bipartite tournament which implies the existence of cycles</u>

An ordinary tournament on n points is acyclic if, and only if, its score sequence consists of the numbers $0, 1, \ldots, n-1$, or equivalently, if, and only if, the sum of the squares of the scores of its points is $\frac{n}{6}(n-1)(2n-1)$. The latter formulation was a problem on the 1958 Putnam Examination [5]. In this section we derive a somewhat similar appearing condition on the score sequence of a bipartite tournament which may indicate whether or not it is acyclic. The price of maintaining simplicity, however, is a loss of generality as the condition to be derived can only show, sometimes, when a bipartite tournament contains cycles.

We have already seen, in §2.10, that if the scores of the points of a bipartite tournament are all strictly positive then the graph contains at least one oriented cycle. Dulmage and Mendelsohn [16] have given a related condition for ordinary unoriented bigraphs.

Theorem 2.11.1. Let there be given an m by n bipartite tournament, $m \geq n$, whose score sequence, arranged in non-decreasing order, is $V = (v_1, v_2, \ldots, v_m)$ and $U = (u_1, u_2, \ldots, u_n)$. If it is acyclic then

$$S_2(m,n) = \sum_{i=1}^{m} v_i^2 + \sum_{j=1}^{n} u_j^2 \geq \frac{n(2n^2+1)}{3} + (m-n)n^2 .$$

For an arbitrary value of m the proof will be by limited induction upon n.

The statement is vacuously true for $n = 0$.

If $n = 1$ there are no cycles possible and $S_2(m,1) = (m-u_1)\cdot 1 + 1 \cdot u_1^2 \geq m$, since $u_1^2 \geq u_1$, which verifies the theorem for this case.

Assume the theorem has been proved, for an arbitrary value of m, for $n = 0, 1, \ldots, k < m$, and consider any acyclic m by $(k+1)$ tournament. Then $u_{k+1} = m-t$, where t is the number of P points whose score is $k+1$, from the discussion of the structure of acyclic bipartite tournaments in general and the hypothesis that the scores were in non-decreasing order. The removal of the point Q_{k+1} and the edges incident upon it leaves an acyclic m by k tournament to which

the induction hypothesis may be applied to imply that

$$S_2(m,k+1)-(m-t)^2 - t[(k+1)^2-k^2] \geq \frac{k(2k^2+1)}{3} + (m-k)\, k^2 \; .$$

But this implies, upon rearranging and using the fact that

$$(m-k-t)^2 \geq m-k-t \; ,$$

that

$$S_2(m,k+1) \geq \frac{(k+1)\,[2(k+1)^2+1]}{3} + (m-k-1)\,(k+1)^2 \; ,$$

which suffices to complete the proof of the theorem by induction.

That equality is possible in the conclusion of this theorem is shown by the m by n bipartite tournament whose score sequence is

$$V = (0, 1, 2, \ldots, n-1, \underbrace{n, \ldots, n}_{m-n \text{ terms}})$$

and

$$U = (1, 2, \ldots, n) \; .$$

The converse of Theorem 2.11.1 is not necessarily valid for it is easy to construct bipartite tournaments whose score sequences satisfy the inequality of the theorem but which contain cycles.

2.12 <u>On the distribution of 4-cycles in random bipartite tournaments</u>

When an ordinary tournament on n points is interpreted as representing an attempt to rank n distinct elements according to some criterion the presence of cycles indicates a degree of inconsistency or lack of transitivity in the ranking process. Significance tests for the degree of consistency present in an attempted ranking of n elements by the method of paired comparisons have been developed in which the

number of cycles of a given type in the corresponding tournament is compared with the number which would be expected, were all the comparisons made at random. The investigation of the complete distribution of cycles in general appears to be difficult. Kendall and Babington-Smith [54], p. 327, express it as follows: "In discussing inconsistences, therefore, it seems best to confine attention to circular triads which, so to speak, constitute the inconsistent elements of the configuration." A circular triad is the same as a cycle of length three, or a 3-cycle, in our terminology. They obtained an expression for the number of 3-cycles a tournament contains in terms of its score sequence and determined the maximum possible number of 3-cycles in any tournament. An independent derivation of the number of 3-cycles in a tournament in terms of the scores by Moser [67] leads to the maximum number of 3-cycles possible in a somewhat more straightforward manner. Moran [66] showed that the distribution of the number of 3-cycles in a tournament, under the hypothesis of randomness, tends to normality as the number of points tends to infinity and obtained the first four moments of the distribution.

Let $C(m,n)$ denote the number of cycles of length four in an m by n tournament. Attention is concentrated on cycles of length four since they are the simplest cycles to study and because if a bipartite tournament has any cycles at all then it has some of length four, as is not difficult to establish. $C(m,n)$ provides, in a sense, a measure of the degree of transitivity, or rather the lack of it, of the relationships indicated by the orientations of the edges, when a bipartite tournament is thought of as representing the outcome of

comparing each member of one population with each member of a second population and making a decision, upon some basis, as to which component of each pair is the preferred one.

The impossibility of obtaining an expression for $C(m,n)$ as a function of the score sequence alone is demonstrated by the following adjacency matrices, A_1 and A_2, which correspond to two bipartite tournaments having the same score sequence but it can be seen that they do not contain the same number of 4-cycles.

$$A_1 = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{vmatrix}, \text{ and } A_2 = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{vmatrix}.$$

In this section we will obtain sharp upper bounds for $C(m,n)$ and prove that under mild restrictions on the relative rates of growth of m and n the distribution of $C(m,n)$ tends to normality as m and n tend to infinity.

For an arbitrary m by n tournament let v_i denote, as usual, the score of P_i, for $i = 1, \ldots, m$. Then

$$(2.12.1) \qquad \sum_{i=1}^{m} v_i(n - v_i)$$

equals the number of ordered pairs of distinct points, (Q_k, Q_ℓ), for which there exists a point, P_i, such that $Q_\ell \to P_i \to Q_k$, counting multiplicities.

Number the ordered pairs of distinct points, (Q_k, Q_ℓ), from 1 to $n(n-1)$ in such a way that if the ordered pair (Q_k, Q_ℓ) corresponds to s, where $s = 1, 2, \ldots, n(n-1)$, then the ordered pair (Q_ℓ, Q_k) corresponds to $n(n-1) - (s-1)$. Let t_s denote the number of times the ordered pair corresponding to s is counted in (2.12.1). Then it follows that

$$(2.12.2) \qquad C(m,n) = \sum_{s=1}^{\binom{n}{2}} t_s \, t_{n(n-1)-(s-1)} \quad .$$

Assume temporarily that $m \equiv n \equiv 0(2)$. We then observe that

$$(2.12.3) \qquad \sum_{s=1}^{n(n-1)} t_s = \sum_{s=1}^{\binom{n}{2}} (t_s + t_{n(n-1)-(s-1)})$$

$$= \sum_{i=1}^{m} v_i(n-v_i) \leq m\, n^2/4 \ .$$

It is now easily seen that an upper bound for $C(m,n)$ in (2.12.2) results by first having

$$(2.12.4) \qquad t_s + t_{n(n-1)-(s-1)} = m$$

for $\frac{n^2}{4}$ values of s and zero for the remaining values and then setting $t_s = t_{n(n-1)-(s-1)}$. This implies

Theorem 2.12.1.

$$C(m,n) \leq \frac{m^2}{4} \cdot \frac{n^2}{4} \quad \text{if} \quad m \equiv n \equiv 0(2) \ .$$

If m is odd then $\frac{m^2}{4}$ is replaced by $\frac{m^2-1}{4}$, and similarly for n, as may be seen by making the appropriate changes in (2.12.3) and (2.12.4).

That the upper bound is sharp for the case $m \equiv n \equiv 0(2)$ is demonstrated by the bipartite tournament in which $P_i \to Q_k$ if, and only if, either $i \leq m/2$ and $k \leq n/2$ or $i > m/2$ and $k > n/2$, otherwise $Q_k \to P_i$. Equally as simple examples suffice to show that the bound is sharp for the other cases also.

In an m by n tournament define a random variable, $S(ij;\, k\ell)$, $i,j = 1, 2, \ldots, m$, $i \neq j$, and $k,\ell = 1, 2, \ldots, n$, $k \neq \ell$, to be 1 or 0 according as to whether the points P_i, P_j, Q_k, and Q_ℓ do or do not form a 4-cycle, respectively. Since only 2 of the 16 equally likely ways of orienting the edges between these points yield a 4-cycle it follows that $E[S(ij;\, k\ell)] = 1/8$. Summing over all suitable pairs, (i,j) and (k,ℓ), implies that

$$E[C(m,n)] = \frac{1}{8}\binom{m}{2}\binom{n}{2}. \tag{2.12.5}$$

Similarly, $E[C^2(m,n)] = E[(\sum S(ij;\, k\ell))^2]$, where the sum is again over the pairs (i,j) and (k,ℓ). In the expansion of this the various types of products involved and the number and expected value of such products are found to be as follows:

	$S(ij;\, k\ell)$	$S(ij;\, k\ell)$	$\binom{m}{2}\binom{n}{2}$	1/8
	$S(ij;\, k\ell)$	$S(ij;\, ks)$	$2n\binom{m}{2}\binom{n-1}{2}$	1/32
	$S(ij;\, k\ell)$	$S(ih;\, k\ell)$	$2m\binom{m-1}{2}\binom{n}{2}$	1/32
	$S(ij;\, k\ell)$	$S(ij;\, rs)$	$\binom{m}{2}\binom{n}{2}\binom{n-2}{2}$	1/64
(2.12.6)	$S(ij;\, k\ell)$	$S(gh;\, k\ell)$	$\binom{m}{2}\binom{m-2}{2}\binom{n}{2}$	1/64
	$S(ij;\, k\ell)$	$S(ih;\, ks)$	$4mn\binom{m-1}{2}\binom{n-1}{2}$	1/64
	$S(ij;\, k\ell)$	$S(ih;\, rs)$	$2m\binom{m-1}{2}\binom{n}{2}\binom{n-2}{2}$	1/64

$S(ij;\ k\ell)$	$S(gh;\ ks)$	$2n\binom{m}{2}\binom{m-2}{2}\binom{n-1}{2}$	$1/64$
$S(ij;\ k\ell)$	$S(gh;\ rs)$	$\binom{m}{2}\binom{m-2}{2}\binom{n}{2}\binom{n-2}{2}$	$1/64$.

Combining these we have, where μ_k denotes the kth moment of $C(m,n)$ about its mean, that

$$(2.12.7) \qquad \mu_2 = E[C^2(m,n)] - E^2[C(m,n)]$$

$$= \frac{1}{2^6}\binom{m}{2}\binom{n}{2}(2m + 2n - 1) .$$

To show that the distribution of $C(m,n)$ tends to normality it suffices (Kendall and Stuart [56], p. 111) to show that, for $h = 1, 2, \ldots,$

$$\text{(i)} \quad \frac{\mu_{2h+1}}{\mu_2^{\frac{1}{2}(2h+1)}} \to 0 \ , \qquad \text{and} \qquad \text{(ii)} \quad \frac{\mu_{2h}}{\mu_2^{h}} \to \frac{(2h)!}{2^h\, h!} \ ,$$

as m and n tend to infinity.

We shall temporarily assume that $n = o(m)$ as m and n tend to infinity.

If $T(ij;\ k\ell) = S(ij;\ k\ell) - 1/8$, then $\mu_{2h+1} =$ $E[(\sum T(ij;\ k\ell))^{2h+1}]$, for $h = 1, 2, \ldots$, where the sum is again over the pairs (i,j) and (k,ℓ). A typical term in the expansion of this, for a fixed value of h , is $T(i_1 j_1;\ k_1 \ell_1) \ldots$ $T(i_{2h+1}\, j_{2h+1};\ k_{2h+1}\, \ell_{2h+1})$. Put all the T's, in this term, which have any values of i, j, k, or ℓ in common with those of the first factor in a class with it. Add to this class any T's, still in this same term, which have any values of i, j, k, or ℓ in common with those of any of the T's already in this class and continue this process as

long as is possible. In a similar fashion form another class starting with, say, the first factor not already included in the first class.

By repeating this process any term may be expressed as the product of classes of products such that no two factors in different classes have any values of i, j, k, or ℓ in common, while for any factor in a class containing more than one factor there is another factor in the same class with which it does have a value of i, j, k, or ℓ in common.

Combining all terms which have similar combinations of (i,j) and (k,ℓ) occurring, as was done in (2.12.6) for the second moment, the number of times terms of a given type appear will be a polynomial in m and n whose largest term is of order equal to the number of distinct values of i and j, and k and ℓ, respectively, that appear in the term.

If any class in a term contains just one factor the expected value of that term will equal the expected value of the product of the remaining factors times the expected value of the single factor, or zero. Also, any term with more than h classes has expected value zero since it is forced to have at least one class containing a single factor.

Restricting ourselves now to terms all of whose classes contain at least two factors we may make the following assertion: If such a term is to have a non-zero expectation, then each factor in a class may include at most one value of i, j, k, or ℓ not included in any other factor in the class. For, if any factor has fewer than three values of i, j, k, and ℓ in common with those of the remaining factors in the

class, then its expectation is independent of the expectations of the remaining factors, as was seen in (2.12.6). As its expectation is zero the expectation of the whole term is consequently zero.

Therefore, for any class of factors in a term the largest number of distinct values of i, j, k, and ℓ that may occur in it is four for the first factor and one new value for each of the remaining factors. Hence, the term containing the largest number of different values of i, j, k, and ℓ with a non-zero expectation will have $h-1$ classes with two factors each and one class of three factors. From the hypothesis on the relative orders of m and n we see that the largest term in μ_{2h+1} is contributed by products of the form

$$\prod_{s=1}^{h-1} [T(i_s j_s; k_s \ell_s)\, T(g_s j_s; k_s \ell_s)]\, T(i_h j_h; k_h \ell_h)\, T(g_h j_h; k_h \ell_h)\, T(f_h j_h; k_h \ell_h) \quad ,$$

where different letters represent different numbers.

Therefore, μ_{2h+1} is a polynomial whose largest term is of degree $3h+1$ in m and $2h$ in n. Hence, for fixed h and m and n tending to infinity in the prescribed manner, it follows that

$$\text{(2.12.8)} \qquad \frac{\mu_{2h+1}}{\mu_2^{\frac{1}{2}(2h+1)}} = O\left(\frac{m^{3h+1}\, n^{2h}}{m^{3/2(2h+1)}\, n^{2h+1}} \right) = O\left(\frac{1}{m^{1/2}\, n} \right) \to 0 \quad .$$

In considering μ_{2h} the same type of argument may be used to assert that its highest ordered terms arise from products of the following type: $\prod_{s=1}^{h} T(i_s j_s;\, k_s \ell_s)\, T(g_s j_s;\, k_s \ell_s)$.

The number of times terms of this type occur is

$$(2h)!\binom{m}{h}\binom{m-h}{2}\binom{n}{2}\ \cdots\ \binom{m-3h+2}{2}\binom{n-2h+2}{2} \sim \frac{m^{3h}\ n^{2h}\ (2h)!}{2^{2h}\ h!} ,$$

retaining only terms of highest order.

The expectation of such a term equals $\{E[T(ij;k\ell)T(gj;k\ell)]\}^h = 1/2^{6h}$, as may be seen by direct considerations. Hence, the leading term of μ_{2h} is $\frac{m^{3h}\ n^{2h}\ (2h)!}{2^{8h}\ h!}$ and that of μ_2 is $m^3n^2/2^7$. Therefore, for fixed h and m and n tending to infinity in the prescribed manner, it follows that

$$(2.12.9) \qquad \frac{\mu_{2h}}{\mu_2^{\ h}} \rightarrow \frac{(2h)!}{2^h\ h!} ,$$

which completes the verification of (i) and (ii) for this case.

For each fixed value of h the only essential difference in the arguments when $n \rightarrow cm$, $0 < c \leq 1$, as m and n tend to infinity, will be the appearance of the additional factor $(1+c)^h$ in both μ_{2h} and $\mu_2^{\ h}$ which clearly leaves (2.12.9) unchanged.

Under the given conditions this completes the proof of Theorem 2.12.2.

$$\Pr\left\{ a < \frac{C(m,n) - \frac{1}{8}\binom{m}{2}\binom{n}{2}}{\frac{1}{8}[\binom{m}{2}\binom{n}{2}(2m+2n-1)]^{\frac{1}{2}}} < b \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-y^2/2}\ dy .$$

2.13 The expected number of cycles in a random bipartite tournament

In spite of the fact that the exact distribution of the number of cycles of a given length in a random tournament does not appear to be known in general it is not difficult to calculate the expected number of cycles of a given length. In this section we treat the corresponding problem for bipartite tournaments.

Let $C_\ell(m,n)$, $\ell = 2, 3, \ldots, n$, denote the number of cycles of length 2ℓ in an m by n tournament, where $m \geq n \geq 2$. Under the hypothesis of randomness it follows that the probability that any set of ℓ distinct P points and ℓ distinct Q points will form a cycle of length 2ℓ is

$$\ell[(\ell-1)!]^2\, 2^{\ell^2-2\ell} / 2^{\ell^2} = \ell[(\ell-1)!]^2 / 2^{2\ell} ,$$

since the points on the cycle can be ordered in $\ell[(\ell-1)!]^2$ ways and there remain $\ell^2-2\ell$ edges each of whose orientation may be chosen in one of two ways. Summing over all such sets of ℓ distinct P points and ℓ distinct Q points gives

$$(2.13.1) \qquad E[C_\ell\,(m,n)] = \binom{m}{\ell}\binom{n}{\ell}\,\ell[(\ell-1)!]^2/2^{2\ell} .$$

For $\ell = 2$ this gives (2.12.5).

A <u>complete</u> (<u>oriented</u>) <u>cycle</u> is one which includes each point of the graph once and only once. A bipartite tournament cannot contain a complete cycle unless $m = n$ in which case

$$(2.13.2) \qquad E[C_m\,(m,m)] = \frac{(m!)^2}{m\,2^{2m}} \sim 2\pi\left(\frac{m}{2e}\right)^{2m} ,$$

from Stirling's formula.

Moser [67] has pointed out that there seem to be no known examples of ordinary tournaments on n points which in general contain as many complete cycles as does the average tournament on n points. For bipartite tournaments this presents no difficulty, however. If m is even consider the m by m bipartite tournament in which $P_i \to Q_j$ if, and only if, both i and j are less than or equal to $m/2$ or both i and j are greater than $m/2$. Otherwise $Q_j \to P_i$. It is easily seen that the number of complete cycles contained in this graph is

$$(2.13.3) \qquad \frac{2}{m}\,[(\tfrac{m}{2})!]^4 \sim 2\,\pi^2\, m\,(\tfrac{m}{2e})^{2m} \sim \pi\, m\, E[C_m(m,m)] \quad .$$

Slight modifications yield asymptotically the same number of complete cycles when m is odd.

A possible conjecture would be that the left member of (2.13.3) is the maximum possible number of complete cycles in an m by m tournament when m is even. The likelihood of this is increased by the fact that it was essentially this same bipartite tournament which contained the maximum possible number of 4-cycles in connection with Theorem 2.12.1.

The expected number of paths of a given length in an m by n tournament can be obtained in the same way as was the expected number of cycles of a given length.

Rédei [80] has shown that every ordinary tournament on n points contains a path of length $n-1$ and, as mentioned in §2.5, an ordinary tournament contains a complete cycle if, and only if, it is irreducible

from results of Camion [6] and Roy [86]. Foulkes proof of this latter statement in [27] is incomplete as the tacit assumption is made that a tournament obtained from an irreducible tournament by the removal of any one point and the edges incident upon it is always irreducible itself, the falsity of which may be demonstrated by examples. That neither the theorems of Rédei nor Camion-Roy hold for m by m bipartite tournaments in general may be shown by examples although the conditions in the latter theorem are clearly necessary. Another condition which is obviously necessary for the existence of a complete cycle in an m by m tournament is that no more than j points in one point set, each of whose score is j, have all the j edges which are oriented away from each of them oriented towards the same j points in the other point set, for $j = 1, 2, \ldots, m$. A conjecture is that this condition along with irreducibility is sufficient to insure the existence of a complete cycle. A slight modification gives a necessary condition for the existence of a path of length $2m-1$ in an m by m tournament.

If one removes the restriction that no point is to be included more than once in a cycle or path then it follows from the definition of matrix multiplication that the total number of cycles of length 2ℓ, $\ell = 2, 3, \ldots$, in an m by n tournament is equal to

$$\frac{1}{\ell} \operatorname{tr} [A \cdot C]^{\ell} = \frac{1}{\ell} \operatorname{tr} [C \cdot A]^{\ell} ,$$

where A and C are the adjacency matrices associated with the tournament as defined in §2.9. The diagonal elements of $[A \cdot C]^{\ell}$ and $[C \cdot A]^{\ell}$ are the number of paths of length 2ℓ beginning and ending

with a given point. The trace is the sum of these numbers and it is necessary to divide by ℓ since each such cycle is counted ℓ times. Various other configurations may be counted in a similar fashion. For corresponding considerations for ordinary directed graphs see e.g. Kemeny, Snell, and Thompson [53], chapter 7, or [11], unit IV - 27.

The problem of determining the number of cycles and paths of a given length, in the sense of our original definition, from the number of cycles and paths where some points may be involved more than once appears to be quite complicated. For ordinary directed graphs this is treated in Harary and Ross, [34] and [37]. These papers contain references to other relevant papers.

2.14 On the number of isolates in a bipartite tournament

By an isolate, in a bipartite tournament, is meant a point whose outdegree is zero. It is impossible for there to be isolates among both the P and Q points of a bipartite tournament simultaneously. In this section the expected value and variance of the number of isolates is derived under the hypothesis of randomness described earlier.

An analogous problem for ordinary directed graphs where the points are interpreted as representing people and the edges represent choices for friends has been treated by Katz [49]. (See also Katz and Powell [52].)

In an m by n tournament in which the outdegrees of the points P_i and Q_j are denoted by v_i and u_j, respectively, let $I_P(m,n; t)$ be the number of isolates among the points of P, where

$\sum_{i=1}^{m} v_i = t$. If A_i denotes the event that P_i , $i = 1, \ldots, m$, is an isolate, then (see e.g. Feller [26], p. 96)

$$\Pr\{I_P(m,n;\ t) = k\} = \sum_{j=0}^{m-k} (-1)^j \binom{k+j}{k} S_{k+j} , \text{ for } k = 0, 1, \ldots, m ,$$

where

$$S_{k+j} = \sum \Pr\{A_{i_1}, A_{i_2}, \ldots , A_{i_{k+j}}\} ,$$

the latter sum being over all distinct combinations of $k+j$ of the events A_i.

But

$$S_{k+j} = \binom{m}{k+j} \binom{n(m-k-j)}{t} / \binom{mn}{t} ,$$

as there are $\binom{m}{k+j}$ ways of selecting the $k+j$ points of P which are to be isolates and $\binom{n(m-k-j)}{t}$ ways of choosing t of the $n(m-k-j)$ edges not incident upon the isolates to be oriented towards a Q point. $\binom{mn}{t}$ is the total number of m by n tournaments containing t edges oriented towards a point of Q.

Therefore,

$$(2.14.1) \qquad \binom{mn}{t} \Pr\{I_P(m,n;\ t) = k\} = \sum_{j=0}^{m-k} (-1)^j \binom{k+j}{k}\binom{m}{k+j}\binom{n(m-k-j)}{t}$$

$$= \binom{m}{k} \sum_{j=0}^{m-k} (-1)^j \binom{m-k}{j}\binom{n(m-k-j)}{t} .$$

In particular,

$$(2.14.2) \qquad \binom{mn}{t} \Pr\{I_P(m,n;t) = 0\} = \sum_{j=0}^{m} (-1)^j \binom{m}{j} \binom{n(m-j)}{t}$$

$$= \begin{cases} 0 \quad , & \text{if} \quad t < m; \\ n^m \quad , & \text{if} \quad t = m; \\ \sum \left[\binom{m}{a_1, a_2, \ldots, a_n} \prod_{i=1}^{n} \binom{n}{i}^{a_i} \right] , & \text{if} \quad m < t \le mn . \end{cases}$$

The sum is over all sets of nonnegative integers, $(a_1, a_2, \ldots, a_n)$, such that

$$1 \cdot a_1 + 2 \cdot a_2 + \ldots + n \cdot a_n = t ,$$

and

$$a_1 + a_2 + \ldots + a_n = m .$$

The proof of the last part is immediate upon observing that if there are a_i P points of outdegree i, for $i = 1, 2, \ldots, n$, the number of ways of selecting these is the multinomial coefficient in the expression and for each point of outdegree i there are $\binom{n}{i}$ ways of selecting i of the Q points such that the edges joining the P point to these points are oriented towards the Q points. The last expression includes the first two as special cases. The identity resulting in these cases is contained in Netto [70], p. 256, but the proof there is by equating coefficients of similar terms in an algebraic identity.

For example, when $m = 3$, $n = 4$, and $t = 5$, (2.14.2) becomes $\binom{3}{0}\binom{4.3}{5} - \binom{3}{1}\binom{4.2}{5} = 624 = \binom{3}{1,2}\binom{4}{1}\binom{4}{2}^2 + \binom{3}{2,1}\binom{4}{1}^2\binom{4}{3}$.

If $m = n$, then

$$(2.14.3)\qquad \Pr\{I_P(m,m;t) = 0\} = \sum_{j=0}^{m} (-1)^j \binom{m}{j} \frac{[m(m-j)]_{(t)}}{(m^2)_{(t)}}$$

$$\leq \sum_{j=0}^{m} (-1)^j \binom{m}{j}(1-j/m)^t \sim e^{-\lambda} \quad ,$$

if $\lambda = m\, e^{-t/m}$ is bounded as m and t tend to infinity. (See Feller [26], p. 94.)

This problem can be considered as being an extension of the classical occupancy problem treated in the last cited reference.

The determination, in this case, of the expected total number of isolates may be obtained from what has been shown and by appealing to symmetry.

When t is not specified let $I_P(m,n)$ denote the number of P points in a given m by n tournament which are isolates. The probability that any particular point of P is an isolate is clearly $1/2^n$. Hence, $I_P(m,n)$ has a binomial distribution and

$$(2.14.4)\qquad E[I_P(m,n)] = m/2^n \quad ,$$

and

$$(2.14.5)\qquad \sigma^2\, [I_P(m,n)] = m(2^n-1)/2^{2n} \quad .$$

This indicates the relationship between the number of isolates in a bipartite tournament and the relative magnitudes of m and n.

If $I_Q(m,n)$ denotes the number of isolates among the Q points of an m by n tournament then $I(m,n) = I_P(m,n) + I_Q(m,n)$ is the total number of isolates in a bipartite tournament. Then

(2.14.6) $E[I(m,n)] = m/2^n + n/2^m$,

and

(2.14.7) $\sigma^2[I(m,n)] = m(2^n-1)/2^{2n} + n(2^m-1)/2^{2m} - 2mn/2^{m+n}$,

from (2.14.4), (2.14.5) symmetry, and the fact that $E[I_P(m,n) \bullet I_Q(m,n)] = 0$. (See e.g. Feller [26], pp. 215-216.)

From Chebyshev's inequality it follows that if m and n satisfy (2.7.1) while tending to infinity it becomes increasingly unlikely that a random bipartite tournament has very many isolates.

The identical distributions apply for the number of P and Q points in a random m by n tournament of outdegrees n and m, respectively.

2.15 On the largest score in a bipartite tournament

The probability that a given P point in a random m by n tournament has score ℓ is clearly $\binom{n}{\ell}/2^n$, for $\ell = 0, 1, \ldots, n$. Hence, the probability that the score of such a point is not more than ℓ is $\sum_{k=0}^{\ell} \binom{n}{k}/2^n$.

Let $M_P(m,n)$ denote the maximum of the scores of the P points in a random m by n tournament. Then

$$(2.15.1) \quad \Pr\{M_P(m,n) = \ell\} = \left\{ \left[\sum_{k=0}^{\ell} \binom{n}{k} \right]^m - \left[\sum_{k=0}^{\ell-1} \binom{n}{k} \right]^m \right\} \cdot 1/2^{mn} ,$$

for $\ell = 0, 1, \ldots, n$, the probability that no score of a point in P

is greater than ℓ minus the probability that all the scores of the points in P are less than ℓ. Therefore,

$$(2.15.2) \qquad E[M_P(m,n)] = \sum_{\ell=0}^{n} \ell \Pr\{M_P(m,n) = \ell\}$$

$$= n - \sum_{\ell=1}^{n} \left[\sum_{k=0}^{\ell-1} (\tfrac{1}{2})^n \binom{n}{k} \right]^m .$$

The inner sum in this expression can be written in terms of an incomplete beta function, tables of values of which are available, e.g. Pearson [72].

Even more difficult would be the obtaining of the expected value of the largest of all the scores in a random bipartite graph, not just restricted to the scores of the points in one subset of points. A solution to the corresponding problem for ordinary tournaments does not appear to have been published, except for some complicated recurrence formulae by David [12].

Let $m = n$. Following Gumbel [33], p. 82, the <u>characteristic largest score</u> among the P points would be that number, k, such that the expected number of P points whose score is greater than or equal to k is one. By definition it would be the number, k, such that

$$\sum_{\ell=0}^{k-1} \binom{n}{\ell} \cdot 1/2^n = 1 - 1/n .$$

Using the normal approximation to the binomial distribution (cf. Feller [26], chapter 7) one finds, as a first approximation, that

$$(2.15.3) \qquad k \approx \frac{n}{2} + \sqrt{\frac{n}{2} \log n} \; .$$

2.16 On the number of locally maximal and locally minimal points in a bipartite tournament

In an m by n tournament a point P_i, whose score is greater than zero, will be said, for lack of a better term, to be locally maximal if, and only if, its score is strictly greater than the score of every point Q_j such that $P_i \to Q_j$. Replacing the final "greater than" by "less than" defines a locally minimal P point. Locally maximal and locally minimal points of Q are defined in an analogous fashion. As defined here the concept of a locally maximal point is not the dual of the concept of a locally minimal point.

In this section we determine the expected number of locally maximal and minimal points in a random bipartite tournament, derive sharp upper bounds for the number of locally maximal points possible in a graph, and count the number of nonisomorphic bipartite tournaments in which this number is attained.

Let $A(m,n;\, P_i)$ be a variable which assumes the value one or zero according as to whether P_i is or is not, respectively, a locally maximal point in a random m by n tournament. Let $B(m,n;\, P_i)$ be defined similarly according as to whether P_i is a locally minimal point. The same variables may be defined for points of Q. Let v_i denote, as usual, the score of P_i, $i = 1, 2, \ldots, m$. The same type of reasoning as was used in §§2.14 and 2.15 implies that

$$\Pr\{A(m,n;P_i) = 1 | v_i = k\} = \left[1 - \sum_{\ell=k}^{m-1} \binom{m-1}{\ell} \cdot 1/2^{m-1} \right]^k ,$$

and

$$\Pr\{B(m,n;P_i) = 1 | v_i = k\} = \left[1 - \sum_{\ell=0}^{k} \binom{m-1}{\ell} \cdot 1/2^{m-1} \right]^k$$

for $k = 1, 2, \ldots, n$, and the left members of both equations equal zero, by definition, when $k = 0$.

Summing over k gives the result that

$$\text{(2.16.1)} \qquad \Pr\{A(m,n;P_i) = 1\} = \sum_{k=1}^{n} \binom{n}{k} \cdot 1/2^n \left[1 - \sum_{\ell=k}^{m-1} \binom{m-1}{\ell} \cdot 1/2^{m-1} \right]^k ,$$

and

$$\Pr\{B(m,n;P_i) = 1\} = \sum_{k=1}^{n} \binom{n}{k} \cdot 1/2^n \left[1 - \sum_{\ell=0}^{k} \binom{m-1}{\ell} \cdot 1/2^{m-1} \right]^k .$$

To obtain corresponding results for Q_j it is necessary only to interchange m and n. It follows that the expected number of locally maximal and locally minimal points in a random m by n tournament is

$$m \Pr\{A(m,n;P_i) = 1\} + n \Pr\{A(m,n;Q_j) = 1\} ,$$

and

$$m \Pr\{B(m,n;P_i) = 1\} + n \Pr\{B(m,n;Q_j) = 1\} , \text{ respectively.}$$

For $m = 3$ and $n = 2$ these values are $139/64$ and $24/64$, respectively. These numbers are in agreement with an actual count of the things involved over the set of sixty-four 3 by 2 tournaments.

Since some point must have a score smaller than the score of any other point it follows that not all of the points in an m by n tournament can be locally maximal simultaneously.

If $m,n > 1$ consider the bipartite tournament in which $P_i \to Q_j$ and $Q_j \to P_1$, for $i = 2, \ldots, m$, and $j = 1, \ldots, n$. If $n = 1$ let $P_i \to Q_1$ for all i. These tournaments contain $m+n-1$ locally maximal elements. This implies the following result:

Theorem 2.16.1. The maximum number of locally maximal points possible in a nontrivial m by n bipartite tournament is $m+n-1$.

We now proceed to determine the number of nonisomorphic m by n bipartite tournaments containing $m+n-1$ locally maximal points, where $m,n \geq 2$.

First we observe that in such a graph there is precisely one point whose score is less than the scores of the remaining points, since there is at least one such point and if there were more than one there would be fewer than $m+n-1$ locally maximal points. Assume temporarily that this point is P_1. Next we observe that among the remaining points it is impossible for a P point and a Q point to have the same score and both be locally maximal points. Hence, those points with the next to the smallest score must, in this case, all be Q points and, furthermore, they must all have score one to be locally maximal since P_1 is the only point having a lower score. This further implies that P_1 must have score zero. Next, those points with the third lowest score in this case, if there are any such points, must all be P points

and each of their scores must equal the number of Q points which had the second lowest scores. In order for these P points to be locally maximal there must be at least two such Q points with score one.

The same type of considerations repeated as often as is necessary suffice to establish the validity of the following assertion: In every nontrivial m by n bipartite tournament containing $m+n-1$ locally maximal points, in which P_1 is the point with minimal score, the points of P and Q may be separated into k and ℓ mutually exclusive and exhaustive non-vacuous classes, where the ith and jth class contain r_i and s_j points, respectively.

$$(2.16.2) \qquad k-\ell = 1 \text{ or } 0 \ , \quad \sum_{i=1}^{k} r_i = m \ , \quad \sum_{j=1}^{\ell} s_j = n \ ,$$

and such that the edges joining the points in the ith class of P points to the points in the first $i-1$ classes of Q points, $i = 2, \ldots, k$, are oriented towards the Q points and all the remaining edges are oriented towards P points. In order for all the points with the exception of P_1 to be locally maximal it must also be the case that

$$(2.16.3) \qquad 1 = r_1 < s_1 < r_1+r_2 < s_1+s_2 < r_1+r_2+r_3 < s_1+s_2+s_3 < \ldots$$
$$< \begin{cases} r_1 + r_2 + \ldots + r_k \ , & \text{if } k = \ell \ ; \\ s_1 + s_2 + \ldots + s_\ell \ , & \text{if } k = \ell+1 \ . \end{cases}$$

Furthermore, each such tournament determines uniquely the r_i's and s_j's, in order, and the converse as well is true up to an

isomorphism. Therefore, the number of nonisomorphic m by n tournaments containing $m+n-1$ locally maximal points such that P_1 is the point with the smallest score is the number of ordered sets of strictly positive integers satisfying (2.16.2) and (2.16.3).

If $k = \ell = r$, let

$$J_{2h} = \sum_{i=1}^{h} s_i \quad \text{and} \quad J_{2h-1} = \sum_{j=1}^{h} r_j , \quad h = 1, \ldots, r .$$

Then the conditions in (2.16.3) may be rewritten as

(2.16.4) $\quad 1 = J_1 < J_2 < J_3 < J_4 < \ldots < J_{2r-1} = m , \quad \text{and} \quad J_{2r} = n .$

By definition it must be that $J_{2(r-1)} \leq \min[m,n] - 1 = \eta$.

In this case it is easily seen that there is a one-to-one correspondence between these sets of numbers and those satisfying (2.16.2) and (2.16.3).

Now let $j_i = J_{i+1} - J_i$, for $i = 1, 2, \ldots, 2(r-1)-1 = 2r-3$. Since $j_i \geq 1$ it follows that $(j_1, j_2, \ldots, j_{2r-3})$ forms a composition of $J_{2(r-1)}-1$ into $2r-3$ parts. (r is at least two in the case we are considering.) Again there is a one-to-one correspondence between the sets consisting of the j_i and the sets of the J_i. But for a fixed value of r and $J_{2(r-1)}$ the number of compositions of $J_{2(r-1)}-1$ into $2r-3$ parts is

$$\binom{J_{2(r-1)}-2}{2(r-2)}$$

as was mentioned in §2.10.

It is not difficult to see that r, in this case, can vary between two and $\min\{[\frac{m+1}{2}], [\frac{n+1}{2}]\} = \alpha$, inclusive, and that for fixed values of r $J_{2(r-1)}$ can vary between $2(r-1)$ and η, inclusive.

Summing over these cases counts the number of m by n tournaments satisfying the given conditions and containing $m+n-1$ locally maximal points. Only slight modifications are necessary to treat the case where $k = \ell+1$. Combining these gives

Theorem 2.16.2. The number of nonisomorphic m by n tournaments, $m,n \geq 2$, containing only one point~~s~~, a P point, which is not a locally maximal point is

$$1 + \sum_{r=2}^{\alpha} \sum_{j=2(r-1)}^{\eta} \binom{j-2}{2(r-2)} + \sum_{r=2}^{\beta} \sum_{j=2r-1}^{\eta} \binom{j-2}{2r-3},$$

where $\beta = \min\{[m/2], [n/2]\}$.

The 1 counts the tournament given as an example in the proof of Theorem 2.16.1 and corresponds to the only graph where $r = 1$ for these cases. When m or n is less than two the problem is trivial. The symmetry of the above expression in m and n indicates that it also represents the number of tournaments in which every point is a locally maximal point except for a single Q point.

Corresponding results about the number of locally minimal points possible in a bipartite tournament appear to be more difficult to obtain. The m by m tournament in which $P_i \to Q_j$ if, and only if, $i = j$, for $i,j = 1, 2, \ldots, m$ contains m locally minimal points for $m \geq 3$ and this can be extended, except for a few small values of m and n, to show that it is possible in general for an m by n

tournament to have as many as $\max[m,n]$ locally minimal points. A conjecture is that this is the maximum number it could have although the only upper bound on the number of locally minimal points possible in an m by n tournament that we have been able to establish is $\max[m,n] + mn/(m+n)$. This arises from the observation that if P_i is the point with lowest score in a bipartite tournament then no point, Q_j, such that $Q_j \to P_i$ can be locally minimal, and the lowest score in an m by n tournament is certainly no more than $mn/(m+n)$.

2.17 A measure of the degree of similarity between different bipartite tournaments

Kendall's τ, [55], provides a measure of the degree of correlation between two different rankings of a set of n objects. The distribution of τ under the hypothesis that all rankings are equally likely has been investigated in order to test the significance of various values of τ in a given situation. In this section we extend some of these considerations to where each element in one set is compared with each element in a second set and a choice is made for each such pair as to which is the preferred element, with respect to some criteria. We shall use the terminology of bipartite tournaments as they provide a model of such situations.

Let there be given two m by n tournaments both of whose point sets are labelled $P = \{P_1, \ldots, P_m\}$ and $Q = \{Q_1, \ldots, Q_n\}$. Let S and N denote, respectively, the number of edges (P_i, Q_j) which are oriented in the same sense and in the opposite sense in the two graphs.

Then $S-N$ may vary between $+mn$ and $-mn$ and

(2.17.1) $$\tau = \frac{S-N}{m\,n} = 1 - \frac{2N}{mn} = \frac{2S}{mn} - 1$$

varies between +1 and -1 , when there is complete agreement and complete disagreement between the two graphs, respectively, providing what may be considered as a <u>coefficient of correlation</u> between the two tournaments.

This can be interpreted as being a special case of the general correlation coefficient, (see Kendall [55], p. 17)

$$\Gamma = \left(\sum a_{ij}\, b_{ij}\right)\left(\sum a_{ij}^2 \cdot \sum b_{ij}^2\right)^{-1/2} ,$$

where a_{ij} is +1 or -1 according as to whether $P_i \to Q_j$ or $Q_j \to P_i$, respectively, in the first tournament and b_{ij} is similarly defined with respect to the second tournament, for $i = 1, \ldots, m$, and $j = 1, \ldots, n$.

Under our hypothesis of randomness the distribution of τ is easy to obtain, it being simply a binomial distribution and a special case of a more general distribution to be treated shortly. The central limit theorem implies that the distribution of $(mn)^{1/2}\,\tau$ tends to the normal distribution with zero mean and unit variance as mn tends to infinity.

Admitting the possibility of a <u>tie</u> between two elements being compared, which may be indicated by the non-existence of an edge joining the corresponding points in the corresponding tournament, introduces complications. In the first place even the concept of correlation

between two sets of rankings one or both of which contain ties admits of various interpretations. For a discussion of this problem see the third chapter of the last cited reference and also Burr [4].

However, for our purposes, the following is as simple an extension as any of the definition of τ to include the case where ties are permitted. If P_i and Q_j are joined by an edge in neither graph the contribution of this pair to S is one, i.e. it counts as an agreement. If they are joined by an edge in only one of the graphs this will contribute nothing to either S or N. In other cases the earlier definition applies.

Let the two m by n tournaments being compared contain t and u ties, respectively, where $t, u = 0, 1, \ldots, mn$. By considering separately the expected contribution of pairs of points which are joined by an edge in neither graph and those which are joined in both graphs the expected value of τ under the hypothesis of randomness is found to be

$$(2.17.2) \qquad E(\tau) = 2[tu/mn + \tfrac{1}{2}(mn-t)(mn-u)/mn] \,/\, mn - 1$$

$$= \frac{3\,tu - (t+u)mn}{(mn)^2} \quad .$$

In determining the variance of τ under these conditions let X_{ij} denote the contribution to S of the pair of points, (P_i, Q_j). As before it follows that

$$E(X_{ij}) = \frac{tu}{(mn)^2} + \frac{1}{2}\,\frac{(mn-t)(mn-u)}{(mn)^2} \quad .$$

Also,

$$\sigma^2(X_{ij}) = E(X_{ij}) \cdot [1 - E(X_{ij})] \quad ,$$

since X_{ij} has a binomial distribution for fixed values of $m, n, t,$ and u.

Furthermore, if not both $i = \ell$ and $j = k$,

$$E(X_{ij}\, X_{\ell k}) = \frac{t(t-1)\, u(u-1)}{[mn(mn-1)]^2} + \frac{tu(mn-t)(mn-u)}{[mn(mn-1)]^2}$$

$$+ \frac{1}{4} \frac{(mn-t)(mn-t-1)(mn-u)(mn-u-1)}{[mn(mn-1)]^2} \quad ,$$

counting separately the contributions from two pairs of points both pairs of which are tied in both graphs, two pairs of points one pair of which is tied in both graphs and the other pair of which contributes one to S by virtue of the fact that the edges joining the pair are oriented the same in both graphs, and those pairs of points both pairs of which contribute to S because of the agreement of the orientations of the edges joining both pairs in the two graphs.

$$\text{cov}\,(X_{ij}\, X_{\ell k}) = E(X_{ij}\, X_{\ell k}) - E^2(X_{ij}) \ ,$$

from the definition.

Therefore, (see e.g. Feller [26], p. 216)

$$(2.17.3) \qquad \sigma^2(\tau) = \frac{4}{(mn)^2}\, \sigma^2(S) = \frac{4}{(mn)^2}\, [mn\, \sigma^2(X_{ij}) + 2\tbinom{mn}{2} \text{cov}(X_{ij}\, X_{\ell k})] \ .$$

If $t = u$, then τ measures the correlation between two

incomplete m by n tournaments, each of which contains $mn-t$ oriented edges. Carrying through the substitutions for this case gives, barring errors, that

$$(2.17.4) \qquad E(\tau) = \frac{t(3t-2mn)}{(mn)^2}$$

and

$$\sigma^2(\tau) = \frac{(mn)^2(mn-1)(mn-2t) + mnt^2(10mn-18t-1) + 9t^4}{(mn)^4\ (mn-1)} .$$

In particular, if $t = u = 0$ this corresponds to the case considered originally.

Kendall has also considered what can be thought of as being a coefficient of agreement between R ordinary tournaments on n points. This can easily be extended to bipartite tournaments where, for simplicity, we restrict ourselves to the case where there are no ties permitted.

Let there be given R , an arbitrary integer greater than one, m by n tournaments all of whose point sets are labelled $P = \{P_1, \ldots, P_m\}$ and $Q = \{Q_1, \ldots, Q_n\}$. Let γ_{ij} denote the number of these tournaments in which $P_i \to Q_j$, for $i = 1, \ldots, m,$ and $j = 1, \ldots, n.$ Then

$$\sum = \sum_{i,j} [\binom{\gamma_{ij}}{2} + \binom{R-\gamma_{ij}}{2}] ,$$

is the total number of pairs of agreements in the orientation of corresponding edges among the R tournaments. A <u>coefficient of agreement</u>, u , can be defined by

$$(2.17.5) \qquad u = \frac{2 \sum}{\binom{R}{2} mn} - 1 ,$$

which is the same as τ when $R = 2$.

From the definition it follows that the minimum value u may assume, if $R > 2$, is $-1/(R-1)$ or $-1/R$, according as to whether R is even or odd, respectively, since it is impossible for there to be complete disagreement as regards the orientation of the edges among more than two tournaments.

Under our hypothesis of randomness it can be shown that

$$(2.17.6) \qquad E(\sum) = \frac{1}{2} \binom{R}{2} mn ,$$

and

$$(2.17.7) \qquad \sigma^2 (\sum) = \frac{1}{4} \binom{R}{2} mn .$$

Also, the statistic

$$\frac{4}{R-2} \left[\sum - \frac{1}{2} mn \binom{R}{2} \frac{R-3}{R-2} \right] , \quad \text{for} \quad R > 2 ,$$

has as its first three moments the first three moments of the χ^2 - distribution with

$$\nu = mn\, R(R-1)/(R-2)^2$$

degrees of freedom.

Except that $\binom{n}{2}$ is replaced by mn the proofs of these statements are almost identical with the proofs of the corresponding statements for ordinary tournaments outlined in Kendall [55], pp. 136-138, and for that reason will be omitted.

CHAPTER III

ON RANDOM BIGRAPHS

3.1 Introduction

Given that a graph has n points and E edges, from these facts alone what can be inferred about the structure of such a graph, i.e. what properties does it possess and what is the nature and number of configurations appearing as subgraphs of such a graph? Many problems of this general type have been considered and in this concluding chapter we treat a few of these and related types for bigraphs.

3.2 On chromatic bigraphs

Let there be given two distinct sets of n points each such that joining each pair of points not both in the same set is an edge which is coloured either red or blue, say, but not both. In such a configuration, called a chromatic bigraph, what is a lower bound for the total number of monochromatic quadrilaterals, i.e. cycles of length four, the four edges of which are of the same colour?

It is easily seen how there may be set up a one-to-one correspondence between chromatic bigraphs and bipartite tournaments. In particular, with every chromatic bigraph may be associated an adjacency matrix of 0's and 1's.

For ordinary chromatic graphs the corresponding problem with respect to triangles was solved by Goodman [32] and later by Sauvé [89].

In that case the number of monochromatic triangles may be expressed as a function of, say, the number of red edges incident on each point, but the two matrices given in §2.12 can be interpreted as corresponding to chromatic bigraphs each of which has the same number of red edges incident on each point but which do not contain the same number of monochromatic quadrilaterals.

For convenience we may equivalently interpret our problem in terms of n by n matrices each of whose entries is either 0 or 1 and where, for any such matrix, $R(n)$ and $B(n)$ denote the number of 2 by 2 minors all of whose entries are 1's and 0's, respectively. We will prove the following

Theorem 3.2.1.

$$R(n) + B(n) \geq \begin{cases} 2u^2(u-1)(4u-3) & , \quad \text{if } n = 4u ; \\ 2u^3(4u-3) & , \quad \text{if } n = 4u+1 ; \\ u\,(2u+1)(4u^2-u-1) & , \quad \text{if } n = 4u+2 ; \\ u^2(2u+1)(4u+3) & , \quad \text{if } n = 4u+3 ; \end{cases}$$

where u is a nonnegative integer.

In any n by n matrix of 0's and 1's , $A = \|a_{ij}\|$, let r_i denote the ith row sum. Then the number of pairs of distinct columns, k and ℓ , for which there exists a row i such that $a_{ik} = a_{i\ell}$ is, counting multiplicities,

$$(3.2.1) \qquad \sum_{i=1}^{n} \left[\binom{r_i}{2} + \binom{n-r_i}{2}\right] .$$

If the $\binom{n}{2}$ pairs of distinct columns have been numbered let

t_v and h_v denote the number of times the vth pair has been counted in (3.2.1) with $a_{ik} = a_{i\ell} = 1$ and $a_{ik} = a_{i\ell} = 0$, respectively.

Then

$$(3.2.2) \qquad \sum_{v=1}^{\binom{n}{2}} t_v = \sum_{i=1}^{n} \binom{r_i}{2} , \qquad \sum_{v=1}^{\binom{n}{2}} h_v = \sum_{i=1}^{n} \binom{n-r_i}{2} ,$$

and

$$(3.2.3) \qquad R(n) + B(n) = \sum_{v=1}^{\binom{n}{2}} \left[\binom{t_v}{2} + \binom{h_v}{2}\right] .$$

Hence, to minimize $R(n) + B(n)$ we should have $\sum_{i=1}^{n} r_i$ as nearly equal to $\sum_{i=1}^{n} (n-r_i)$ as possible, the r_i's as nearly equal to each other as possible, the t_v's as nearly equal to each other as possible, and similarly for the h_v's.

When $n = 4u$ the smallest possible value for $R(n) + B(n)$ would occur when all $r_i = 2u$ and t_v was equal to u for $2u(3u-1)$ values of v and $u-1$ for the remaining $2u^2$ values of v , since

$$u[2u(3u-1)] + (u-1)(2u^2) = 4u \binom{2u}{2} = \sum_{i=1}^{n} \binom{r_i}{2} ,$$

and similarly for the h_v's. Hence, for any matrix in this case

$$(3.2.4) \qquad R(n) + B(n) \geq 2\left[2u(3u-1) \binom{u}{2} + 2u^2 \binom{u-1}{2}\right]$$

$$= 2u^2(u-1)(4u-3) .$$

In the same way it is found that if $n = 4u+2$ the minimum would occur when all $r_i = 2u+1$, t_v was equal to u for $(2u+1)(3u+1)$ values of v and $u+1$ for the remaining $u(2u+1)$ values, and similarly for the h_v's. In this case

$$(3.2.5) \qquad R(n) + B(n) \geq 2[(2u+1)(3u+1)\binom{u}{2} + u(2u+1)\binom{u+1}{2}]$$

$$= u(2u+1)(4u^2-u-1) .$$

When $n = 4u+3$ it is obviously impossible to have $\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} (n-r_i)$. The nearest to equality that these two could come would be if r_i, say, was $2u+2$ for $2u+2$ values of i and $2u+1$ for the remaining $2u+1$ values. The other conditions, in this case, would require that t_v be equal to $u+1$ for $2(u+1)(2u+1)$ values of v and u for the remaining $(2u+1)^2$ values and that h_v be equal to $u+1$ for $(2u+1)^2$ values of v and u for the other values. Hence, in this case,

$$(3.2.6) \qquad R(n) + B(n) \geq (2u+1)(4u+3)\,[\binom{u+1}{2} + \binom{u}{2}] = u^2(2u+1)(4u+3) .$$

When $n = 4u+1$ the same sort of reasoning implies that the minimum value of $R(n) + B(n)$ would occur when, say, r_i was equal to $2u+1$ for $2u+1$ values of i and $2u$ for the remaining values; t_v was equal to u for $8u^2+u$ values of v and $u+1$ for the other u values; and when h_v was equal to u for $8u^2+u$ values of v and $u-1$ for the remaining u values. As a lower bound this gives

$$(3.2.7) \qquad R(n) + B(n) \geq 2(8u^2+u)\binom{u}{2} + u[\binom{u+1}{2} + \binom{u-1}{2}]$$

$$= u[(4u+1)(u-1)(2u)+1] .$$

But if $u > 0$ it is impossible for equality ever to hold here. To see this we notice that, by symmetry, a matrix for which $R(n) + B(n)$ is to attain the minimum in (3.2.7) must have, in this case, $2u$ columns whose column sum is $2u$. For each of the $\binom{2u}{2}$ pairs of such columns t_v is u or $u+1$ and h_v is $u-1$ or u. Upon attempting to construct two columns in a $4u+1$ by $4u+1$ matrix each containing $2u$ ones and $2u+1$ zeros it is not difficult to see that none of these combinations of t_v and h_v are possible, and that the next best combinations from the point of view of minimizing the sum is (3.2.3) are $t_v = u$ and $h_v = u+1$ or $t_v = u-1$ and $h_v = u$. For each $h_v = u+1$, for some of these pairs of columns, there must be an extra $h_v = u-1$ in order to preserve equality in (3.2.2) and a similar remark can be made with respect to those t_v which equal $u-1$. Clearly there were enough t_v's and h_v's originally set equal to u to permit all the necessary changes. Since $\binom{u+1}{2} + \binom{u-1}{2} = 2\binom{u}{2} + 1$, we see that whichever of these changes we make we are forced to increase the sum in (3.2.3) by one for each of the $\binom{2u}{2}$ pairs of distinct columns whose column sum is $2u$.

Adding this to the lower bound obtained previously in (3.2.7) gives the result, when $n = 4u+1$, that

$$(3.2.8) \qquad R(n) + B(n) \geq 2u^3 (4u-3) ,$$

which completes the proof of Theorem 3.2.1.

It would be of interest to be able to give a general construction which would show that the bounds given in this theorem are sharp or to give larger bounds if they are not. That they are not far from being

best possible is implied by the fact that if each of the n^2 edges is equally as likely to be coloured red as it is to be coloured blue then the expected value of $R(n) + B(n)$ is $\frac{1}{8}\binom{n}{2}^2$, which is only slightly larger than the lower bounds derived in the theorem.

The following 9 by 9 matrix

$$A = \begin{vmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \end{vmatrix} ,$$

contains forty-four 2 by 2 minors each of whose entries is 1 and thirty-six 2 by 2 minors each of whose entries is 0 which shows that equality may occur in the bounds given for $R(n) + B(n)$ when $n = 9$. Other examples suffice to show that the bounds are sharp for $n \leq 9$.

The problem treated here as well as the argument used admit of obvious generalizations.

Related to this problem is one due to Zarankiewicz [93] which may be expressed as follows: Find a function, $A(m,n; k,\ell)$, such that every m by n matrix of 0's and 1's which contains more than $A(m,n; k,\ell)$ 1's contains a k by ℓ minor all of whose entries are 1's. From the first parts of (3.2.1) to (3.2.3) it follows that

if an n by n matrix of 0's and 1's contains A 1's then the number of 2 by 2 minors all of whose entries are 1's contained in the matrix is at least

$$(3.2.9) \qquad \binom{n}{2}\binom{n\binom{A/n}{2}\Big/\binom{n}{2}}{2} .$$

This can be solved for A in terms of n to determine when it is positive. Doing so gives the result that

$$(3.2.10) \qquad A(n,n;\, 2,2) \;\leq\; \frac{n}{2}\,(1+\sqrt{4n-3}) .$$

This has been shown by Rieman [82].

More generally the same type of reasoning can be extended to show that

$$(3.2.11) \qquad A(m,n;\, k,\ell) \;\leq\; (k-1)n + (\ell-1)^{1/k}\, n^{1-1/k}\, m ,$$

which has been derived, using a somewhat different approach, by Hyltén-Cavallius [48]. Other partial results are contained in papers by Kövari, Sós, and Turán [58] and Čulik [10].

Another related problem has to do with <u>unit graphs</u>, by which is meant a graph in the Euclidean plane such that every pair of points which are adjacent are a unit distance apart, measured in the ordinary sense. Moser [67] has shown, among other things, that if an ordinary graph on n points with E edges is an unit graph, then

$$(3.2.12) \qquad E \;\leq\; \frac{n}{4}\,(1+\sqrt{8n+7}) ,$$

improving an earlier result due to Erdös [18].

For m by n unit bigraphs a similar approach may be used although the result which is thus obtained and (3.2.12) are both quite likely far from being best possible. The critical step is to observe that for each of the $\binom{m}{2}$ pairs of distinct P points it is impossible for the points of such a pair to be both joined to more than two Q points at a unit distance from each of them. This provides an upper bound for the number of Q points such that there are two P points at a unit distance from it which are adjacent to it, counting multiplicities. A lower bound for this same quantity can be obtained in terms of the number of edges the graph contains by the means used to obtain (3.2.1) and (3.2.2). Solving the resulting inequality, the details of which we omit, gives the result that if an m by n unit bigraph contains E edges then

$$(3.2.13) \qquad E \leq \frac{n + \sqrt{n^2 + 8mn\,(m-1)}}{2} .$$

m and n may be interchanged in this inequality since the original problem is symmetric in m and n.

3.3 On the largest monochromatic subgraph of a chromatic bigraph

A set-theoretical theorem due to Ramsey [75] implies, among other things, the existence of a function, $N(k,\ell)$, where k and ℓ are positive integers, such that any ordinary chromatic graph with at least $N(k,\ell)$ points contains either a complete subgraph on k points all of whose edges are red or a complete subgraph on ℓ points all of whose edges are blue. Various bounds have been derived for $N(k,\ell)$. See, for example, Erdös and Szekeres [17], Erdös [19], [20], [21],

and [24], and Greenwood and Gleason [31] for some of these and related results. In this section a special case of a corresponding problem for chromatic bigraphs will be considered.

As in previous sections it will be more convenient to express the problem in terms of matrices. Let $k(m,n)$ be the largest number, k, such that every m by n matrix of 0's and 1's contains a k by k minor all of whose entries are the same. With no loss of generality we may assume that $m \geq n$.

Lemma 3.3.1. $k(5,5) = 2$.

Each row in a 5 by 5 matrix of 0's and 1's contains at least three identical entries. Hence, there are at least three rows each of which contains at least three 1's, say. Essentially the only way of filling out two rows of such a matrix, each containing three 1's, without having a 2 by 2 minor involving these rows all of whose entries are identical is, say, if

$$b_{11} = b_{12} = b_{13} = 1, \quad b_{14} = b_{15} = 0,$$

$$b_{21} = b_{22} = 0, \quad b_{23} = b_{24} = b_{25} = 1,$$

where b_{ij} denotes the element in the ith row and jth column of the matrix. But it is not difficult to see that it is impossible to add a third row containing at least three 1's without having at least one 2 by 2 minor all of whose entries are identical. This shows that $k(5,5) \geq 2$, which also is a consequence of Theorem 3.2.1.

That $k(5,5) \leq 2$ is shown by the following matrix:

$$A = \begin{vmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{vmatrix} \quad .$$

This lemma could be thought of as being the bipartite analogue of the statement that of every six people at least three are either mutual acquaintances or mutual strangers. (See e.g. Goodman [32].)

We now restrict ourselves to the case that $m,n \geq 5$ and shall further assume that $\log m = o(n^{\frac{1}{2}})$ as m and n tend to infinity.

If every m by n matrix of 0's and 1's contains a k by k minor of identical elements then the number of ways of forming such a minor times the number of ways of choosing all the remaining entries of the matrix is at least as great as the total number of m by n matrices of 0's and 1's. That is

$$2\binom{m}{k}\binom{n}{k} 2^{mn-k^2} \geq 2^{mn} \quad ,$$

or

$$(mn)^k \geq 2^{k^2-1} (k!)^2 > 2^{k^2} \quad ,$$

since $k \geq 2$ for $m,n \geq 5$.

This implies that

(3.3.1) $\qquad k(m,n) < \log_2(mn)$.

Next consider an m by n matrix of 0's and 1's which contains no $k+1$ by $k+1$ minor all of whose entries are identical. Such a matrix

exists from the definition of $k(m,n)$. With no loss of generality we may assume that this matrix contains at least as many 1's as it does 0's. Hence

$$\sum_{i=1}^{m} r_i \geq mn/2 \quad ,$$

where r_i denotes the ith row sum. The number of (k+1)-tuples of 1's in the ith row is $\binom{r_i}{k+1}$. Then it must be that

$$\sum_{i=1}^{m} \binom{r_i}{k+1} \leq k\binom{n}{k+1} \quad ,$$

for otherwise there would be at least one (k+1)-tuple of columns such that $k+1$ or more rows all had 1's in these columns, forming a $k+1$ by $k+1$ minor of 1's contrary to the assumption. The sum in the left member of this inequality is minimized when all the r_i are equal to $n/2$. Therefore,

$$m \binom{n/2}{k+1} \leq k \binom{n}{k+1} \quad .$$

From (3.3.1) and the assumption that $\log m = o(n^{\frac{1}{2}})$ it follows that $k = o(n^{\frac{1}{2}})$ as m and n tend to infinity. This being the case it follows from Stirling's formula that for sufficiently large m and n the last inequality is equivalent to

$$m \frac{(n/2)^{k+1}}{(k+1)!} \leq k \frac{n^{k+1}}{(k+1)!} \quad ,$$

or

$$(3.3.2) \qquad m \leq k\, 2^{k+1} \quad .$$

Combining (3.3.1) and (3.3.2) gives

Theorem 3.3.1. For every $\epsilon > 0$ the following inequality holds for sufficiently large m and n if $\log m = o(n^{\frac{1}{2}})$ as m and n tend to infinity.

$$(1-\epsilon) \log_2 m < k(m,n) < \log_2(mn) .$$

Various generalizations may be treated in a similar fashion.

3.4 Threshold functions for certain subgraphs of bigraphs

In a series of papers, [22], [23], and [25], Erdős and Rényi have treated several questions dealing with the evolution of random graphs. With respect to some property a graph may possess, a typical problem is to determine how many edges a graph should have in order for it to be reasonably certain, in some sense, that it should possess the given property. Presumably, for each such theorem there should exist a corresponding result for bigraphs. In this section we indicate how a typical argument appearing in [23] can be extended to treat bigraphs. The result to be derived here is the bipartite analogue of Theorem 1 in the last cited paper and the terminology used there will be retained insofar as is possible.

An m by n bigraph with t edges, $0 \leq t \leq mn$, will be denoted by $\Gamma(m,n;\, t)$ and it is understood that each of the $\binom{mn}{t}$ such configurations is equally as likely to be the graph under consideration.

We shall denote by $P_{m,n;t}(A)$ the probability that the random graph $\Gamma(m,n;t)$ possesses a given property A. If, for a given property A , there exists a function, $A(m,n)$, such that

$$\lim_{m,n\to\infty} P_{m,n;t}(A) = \begin{cases} 0 \ , & \text{if} \ \lim\limits_{m,n\to\infty} \dfrac{t(m,n)}{A(m,n)} = 0 \ , \\ \\ 1 \ , & \text{if} \ \lim\limits_{m,n\to\infty} \dfrac{t(m,n)}{A(m,n)} = +\infty \ , \end{cases}$$

where $t = t(m,n)$, then $A(m,n)$ will be called a threshold function of the property A.

First it may be observed that if $t = o(\sqrt{mn/(m+n)})$ it is very likely that the graph consists of isolated points and isolated edges having no points in common. For, the probability that any two distinct edges of $\Gamma(m,n;\ t)$ have an endpoint in common is easily seen to be

$$1 - \frac{\binom{m}{t}\binom{n}{t} t!}{\binom{mn}{t}} \sim 1 - \exp\left[-t^2 \left(\frac{m+n-1}{2mn}\right) - t^3\left(\frac{m^2+n^2-1}{6\,m^2\,n^2}\right)\right] \ ,$$

$$= O[t^2(m+n)/mn] \ ,$$

using the fact, a consequence of Stirling's formula, that

$$\binom{N}{k} \sim \frac{N^k}{k!} \exp[-k^2/2N - k^3/6N^2] \ , \quad \text{if} \quad k = o(N^{3/4}) \ .$$

The P and Q degrees of $\Gamma(m,n;\ t)$ are defined to be t/m and t/n , respectively. These are simply the average degrees of the points in P and Q , respectively. $\Gamma(m,n;\ t)$ will be said to be balanced if, and only if, the P and Q degrees of every proper subgraph of $\Gamma(m,n;\ t)$ are less than or equal to the P and Q degrees, respectively, of $\Gamma(m,n;\ t)$ itself.

If $a + b - 1 \leq \ell \leq ab$, $a,b \geq 1$, let $\beta = \beta(a,b;\ \ell)$ denote an arbitrary non-empty set of connected and balanced a by b bigraphs with ℓ edges. Let there be $B = B(a,b;\ \ell)$ graphs in β. We will prove the following result:

Theorem 3.4.1. The threshold function for the property that the random graph $\Gamma(m,n;\ t)$ should contain at least one subgraph isomorphic with some graph in $\beta(a,b;\ \ell)$ is $m^{1-a/\ell}\ n^{1-b/\ell}$.

Let $P(\beta)$ denote the probability described in the theorem. Then

$$(3.4.1) \qquad P(\beta) \leq \binom{m}{a}\binom{n}{b} \frac{B\binom{mn-\ell}{t-\ell}}{\binom{mn}{t}} = O(t^{\ell}\ m^{a-\ell}\ n^{b-\ell}),$$

the number of ways of choosing the a P points and b Q points upon which may be formed one of B graphs isomorphic with some graph in β times the number of ways of distributing the remaining edges. The inequality arises from the fact that possibly some graphs are counted more than once in this process. Since the last expression is $o(1)$ if $t = o\,(m^{1-a/\ell}\ n^{1-b/\ell})$ this suffices to establish part of the theorem by the definition of a threshold function.

Let $G = G^{m,n}_{a,b;\ell}$ denote the set of all subgraphs of the complete m by n bigraph which are isomorphic with some element of $\beta(a,b;\ \ell)$. With any graph S of this set associate a random variable $\mu(S)$ which assumes the value 1 or 0 according as to whether S is or is not a subgraph of $\Gamma(m,n;\ t)$. Then

$$(3.4.2) \qquad E\left[\sum_{S\in G} \mu(S)\right] = \sum_{S\in G} E\,[\mu(S)]$$

$$= \binom{m}{a}\binom{n}{b} B \frac{\binom{mn-\ell}{t-\ell}}{\binom{mn}{t}} \sim \frac{B}{a!\ b!} \cdot \frac{t^{\ell}}{m^{\ell-a}\ n^{\ell-b}} ,$$

for essentially the same reasons as before.

If S_1 and S_2 are both in G and have no edges in common then clearly

$$E[\mu(S_1)\ \mu(S_2)] = \frac{\binom{mn-2\ell}{t-2\ell}}{\binom{mn}{t}} ,$$

and the number of pairs of such graphs is not more than

$$[\binom{m}{a}\ \binom{n}{b}\ B]^2 .$$

On the other hand if S_1 and S_2 contain exactly r , $1 \leq r \leq \ell-1$, edges in common and u and v common P and Q points, respectively, then it follows that

$$E[\mu(S_1)\ \mu(S_2)] = \frac{\binom{mn-2\ell+r}{t-2\ell+r}}{\binom{mn}{t}} = O[(t/mn)^{2\ell-r}] .$$

Applying to the intersection of S_1 and S_2 the hypothesis that the graphs in β were balanced implies that $u \geq a\ r/\ell$ and $v \geq b\ r/\ell$. Hence, the total number of such pairs of subgraphs, S_1 and S_2 , is not more than

$$\binom{m}{a}\binom{n}{b} B^2 \cdot \sum_{u \geq a\ r/\ell}^{a} \binom{a}{u}\binom{m-a}{a-u} \cdot \sum_{v \geq b\ r/\ell}^{b} \binom{b}{v}\binom{n-b}{b-v}$$

$$= O[m^{a(2-r/\ell)}\ n^{b(2-r/\ell)}] ,$$

where the first two factors represent the number of ways of choosing which points are to form part of a subgraph isomorphic with some element of β and the sums represent the number of ways of choosing the points which are to form part of a second subgraph, having some points in common with the first, isomorphic with another element of β. B^2 is an upper bound to the number of ways of choosing which two elements of β are to be involved.

Therefore,

$$(3.4.3) \qquad E\left[\left(\sum_{S\in G} \mu(S)\right)^2\right] \leq \sum_{S\in G} E[\mu(S)]$$

$$+ \left[\binom{m}{a}\binom{n}{b}B\right]^2 \frac{\binom{mn-2\ell}{t-2\ell}}{\binom{mn}{t}} + 0\left[\left(\frac{t^\ell}{m^{\ell-a}n^{\ell-b}}\right)^2 \sum_{r=1}^{\ell-1} \left(\frac{m^{1-a/\ell}n^{1-b/\ell}}{t}\right)^r\right]$$

multiplying bounds for the number of terms of various types in the expansion by the expected value of such terms, as already determined.

But

$$\frac{\binom{mn-2\ell}{t-2\ell}}{\binom{mn}{t}} \leq \frac{\binom{mn-\ell}{t-\ell}^2}{\binom{mn}{t}^2} ,$$

and supposing that

$$\frac{t}{m^{1-a/\ell}n^{1-b/\ell}} = \omega \to +\infty$$

gives from (3.4.2) and (3.4.3) the result that

$$(3.4.4) \qquad \sigma^2\left[\sum_{S\in G} \mu(S)\right] = 0\left\{\left[\sum_{S\in G} E(\mu(S))\right]^2 \cdot 1/\omega\right\} .$$

Chebyshev's inequality may be used to imply that

$$P_{m,n;t}\left\{\left|\sum_{S\in G}\mu(S)-\sum_{S\in G}E(\mu(S))\right|\geq\frac{1}{2}\sum_{S\in G}E(\mu(S))\right\}=O(1/\omega)\ ,$$

or that

$$(3.4.5)\qquad P_{m,n;t}\left(\sum_{S\in G}\mu(S)\leq\frac{1}{2}\sum_{S\in G}E(\mu(S))\right)=O(1/\omega)\ .$$

But from (3.4.2) we see that as $\omega\to\infty$ so also does $\sum_{S\in G} E(\mu(S))$ which certainly implies from (3.4.5) that the probability that $\Gamma(m,n;\ t)$ contains at least one subgraph isomorphic with some element of $\beta(a,b;\ \ell)$ tends to one under the given assumptions. This completes the proof of Theorem 3.4.1 by the definition of a threshold function.

The following results are immediate consequences upon observing that the subgraphs involved satisfy the hypothesis of the theorem.

Corollary 3.4.1. The threshold function for the property that the random graph, $\Gamma(m,n;\ t)$, contains a subgraph which is a complete a by b bigraph is $m^{1-1/b}\ n^{1-1/a}$.

Corollary 3.4.2. The threshold function for the property that the random graph, $\Gamma(m,n;\ t)$, contains a subgraph which is a cycle of length $2k$, $k\geq 2$, is $(mn)^{1/2}$.

Corollary 3.4.3. The threshold function for the property that the random graph, $\Gamma(m,n;\ t)$, contains a subgraph which is a path of length $2k-1$ is $(mn)^{\frac{k-1}{2k-1}}$; the corresponding function if the path is of length $2k$ with $k+1$ P points is $(mn)^{1/2}\ m^{-1/2k}$.

Corollary 3.4.4. The threshold function for the property that the random graph, $\Gamma(m,n;\ t)$, contains a subgraph which is a tree with a P points and b Q points is $m^{\frac{b-1}{a+b-1}} n^{\frac{a-1}{a+b-1}}$.

For an account of other problems dealing with random graphs which could be considered for bipartite graphs also but which will not be included here see Katz [50], Austin, Fagen, Penney, and Riordan [1] and Harris [47].

3.5 On the number of dyads in directed bigraphs

In any directed graph two distinct points are said to form a dyad if they are joined by two edges which are directed in opposite senses. Interpreting the points as people and the directed edges as friendship choices, say, then a dyad represents a mutual choice between two people. Katz and Wilson [51] have investigated the number of dyads in such a setting under a suitable hypothesis of randomness. One obvious extension of this problem is to assume that each person belongs to one of two classes, but not both, and restricts his choices to members of the other class. The tacit assumption is that no person is permitted to choose another person more than once.

In the terminology of graphs the problem becomes the following: Let there be given a directed m by n bigraph in which the points P_i and Q_j have outdegrees g_i and b_j, respectively, where $0 \leq g_i \leq n$ and $0 \leq b_j \leq m$, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. We will derive expressions for the expected value and variance of D, the number of dyads, i.e. pairs of points, P_i and Q_j, such that $P_i \to Q_j$ and $Q_j \to P_i$, in the graph, under the hypothesis that all

such configurations possible are equally likely and that the arrangement of the edges issuing from any one point is independent of the arrangement of the edges issuing from any other point.

Let X_{ij}, $i = 1, \ldots, m$ and $j = 1, \ldots, n$, denote a random variable which assumes the value 1 or 0 according as to whether the points P_i and Q_j do, or do not, form a dyad. Also, let

$$G_h = \sum_{i=1}^{m} g_i^h$$ for $h = 1, 2,$ the power sum symmetric function, and

let B_h be similarly defined. In what follows the range of summation for various indices will be omitted when it is clear from the context what it should be.

$$\text{(3.5.1)} \qquad E(D) = \sum E(X_{ij}) = \sum \frac{g_i\, b_j}{n\, m} = \frac{G_1\, B_1}{n\, m} \quad ,$$

from the hypothesis and definitions of the symbols involved.

The hypothesis of independence may be used to imply that

$$E(X_{ij}\, X_{k\ell}) = \frac{g_i\, b_j}{n\, m} \cdot \frac{g_k\, b_\ell}{n\, m} \, , \quad \text{if} \quad i \neq k \quad \text{and} \quad j \neq \ell \, .$$

Furthermore, it is easily seen that

$$E(X_{ij}\, X_{i\ell}) = \frac{\binom{g_i}{2}}{\binom{n}{2}} \cdot \frac{b_j\, b_\ell}{m^2} \; ; \quad \text{if} \quad j \neq \ell \quad ,$$

and, by symmetry,

$$E(X_{ij}\,X_{kj}) = \frac{g_i\,g_k}{n^2} \cdot \frac{\binom{b_j}{2}}{\binom{m}{2}} \quad , \quad \text{if} \quad i \neq k \; .$$

Finally,

$$E(X_{ij}\,X_{ij}) = \frac{g_i\,b_j}{n\,m} \; .$$

Therefore,

$$(3.5.2) \qquad E(D^2) = E\left[\left(\sum X_{ij}\right)^2\right] = \sum \frac{g_i\,b_j}{n\,m}$$

$$+ 2 \sum_{\substack{i \neq k \\ j \neq \ell}} \frac{g_i\,g_k\,b_j\,b_\ell}{(mn)^2} + 2 \sum_{j \neq \ell} \frac{g_i(g_i-1)\,b_j\,b_\ell}{n(n-1)\;m^2}$$

$$+ 2 \sum_{i \neq k} \frac{b_j(b_j-1)\,g_i\,g_k}{m(m-1)\;n^2} = \frac{G_1\,B_1}{n\,m}$$

$$+ \frac{1}{2}\,\frac{(G_1^2-G_2)(B_1^2-B_2)}{(nm)^2} + \frac{(G_2-G_1)(B_1^2-B_2)}{n(n-1)\;m^2} + \frac{(B_2-B_1)(G_1^2-G_2)}{m(m-1)\;n^2} \; ,$$

upon expanding and collecting terms in the original expression and taking the expected value of each term.

For given values of the g_i's and b_j's the various expressions appearing may be computed with no difficulty which enables the variance of D to be determined by using

$$(3.5.3) \qquad \sigma^2\,(D) = E(D^2) - [E(D)]^2 \; .$$

When all $g_i = g$ and all $b_j = b$ some simplification occurs;

in particular when $g = b = 1$ (3.5.1) and (3.5.3) become

$$(3.5.4) \qquad E(D) = 1 ,$$

and

$$(3.5.5) \qquad \sigma^2(D) = \frac{(m-1)(n-1)}{m\ n} .$$

In this last case, where the outdegree of every point is one, let $D(m,n)$ be the number of dyads in a directed m by n graph under the same assumptions about randomness as given earlier. Let $m \geq n$. It is an immediate consequence of the principle of cross-classification (cf. Feller [26], p. 96) that

$$(3.5.6) \qquad Pr(D(m,n)=k) = \sum_{j=0}^{n-k} (-1)^j \binom{k+j}{k}\binom{m}{k+j}\binom{n}{k+j}(k+j)! \frac{n^{m-k-j}\ m^{n-k-j}}{n^m\ m^n}$$

$$= \frac{1}{k!} \frac{m_{(k)}\ n_{(k)}}{m^k\ n^k} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \frac{(m-k)_{(j)}\ (n-k)_{(j)}}{m^j\ n^j} ,$$

for $k = 0, 1, \ldots , n$.

In particular

$$(3.5.7) \qquad Pr(D(m,n)=0) = \sum_{j=0}^{n} \frac{(-1)^j}{j!} \frac{m_{(j)}\ n_{(j)}}{m^j\ n^j} ,$$

which provides another generalization of the "problême des recontres." (See e.g. Riordan [84], p. 57.)

BIBLIOGRAPHY

[1] Austin, T. L., Fagen, R. E., Penney, W. F., and Riordan, J., "The number of components in random linear graphs," Annals of Math. Statistics 30 (1959) 747-754.

[2] Austin, T. L., "The enumeration of point labelled chromatic graphs and trees," Canadian Journal of Math. 12 (1960) 535-545.

[3] Berge, C., "Théorie des graphes et ses applications," Paris, Dunod, 1958.

[4] Burr, E. J., "The distribution of Kendall's score S for a pair of tied rankings," Biometrika 47 (1960) 151-171.

[5] Bush, L. E., "The William Lowell Putnam Mathematical Competition," American Math. Monthly 68 (1961) 18-33.

[6] Camion, P., "Chemins et circuits hamiltoniens des graphes complets," C. R. Acad. Sci. Paris 249 (1959) 2151-2152.

[7] Cayley, A., "On the theory of analytical forms called trees," Philosophical Magazine 13 (1857) 172-176. See also "The collected mathematical papers of A. Cayley," Cambridge, 3 (1890) 242-246.

[8] Cayley, A., "A theorem on trees," Quarterly Journal of Pure and Applied Math. 23 (1889) 376-378. See also "The collected mathematical papers of A. Cayley," Cambridge, 13 (1897) 26-28.

[9] Clarke, L. E., "On Cayley's formula for counting trees," Journal London Math. Soc. 33 (1958) 471-475.

[10] Čulík, K., "Teilweise Lösung eines verallgemeinerten Problems von K. Zarankiewicz," Annales Polonici Mathematici 3 (1956) 165-168.

[11] Dartmouth College Writing Group, "Modern Mathematical Methods and Models," vol. 2, Math. Association of America, (1958).

[12] David, H. A., "Tournaments and paired comparisons," Biometrika 46 (1959) 139-149.

[13] Davis, R. L., "The number of structures of finite relations," Proc. American Math. Soc. 4 (1953) 486-495.

[14] Davis, R. L., "Structures of dominance relations," Bull. Math. Biophysics 16 (1954) 131-140.

[15] Dulmage, A. L., and Mendelsohn, N. S., "Coverings of bipartite graphs," Canadian Journal of Math. 10 (1958) 517-534.

[16] Dulmage, A. L., and Mendelsohn, N. S., "The term and stochastic rank of a matrix," Canadian Journal of Math. 11 (1959) 269-279.

[17] Erdös, P. and Szekeres, G., "A combinatorial problem in geometry," Compositio Mathematica 2 (1935) 463-470.

[18] Erdös, P., "On sets of distances of n points," American Math. Monthly 53 (1946) 248-250.

[19] Erdös, P., "Some remarks on the theory of graphs," Bull. American Math. Soc. 53 (1947) 292-294.

[20] Erdös, P., "Remarks on a theorem of Ramsey," Bull. Research Council of Israel 7F (1957) 21-24.

[21] Erdös, P., "Graph theory and probability," Canadian Journal of Math. 11 (1959) 34-38.

[22] Erdös, P. and Rényi, A., "On random graphs I ," Publicationes Mathematicae (Debrecen) 6 (1959) 290-297.

[23] Erdös, P. and Rényi, A., "On the evolution of random graphs," Publications of the Math. Inst. of the Hungarian Acad. of Sci. 5 (1960) 17-61.

[24] Erdös, P., "Graph theory and probability, II," Canadian Journal of Math. 13 (1961) 346-352.

[25] Erdös, P. and Rényi, A., "On the strength of connectedness of a random linear graph," Acta Mathematica Academiae Scientiarum Hungaricae 12 (1961) 261-267.

[26] Feller, W., "An Introduction to Probability Theory and its Applications," New York, John Wiley and Sons, Inc., 1957.

[27] Foulkes, J. D. "Directed graphs and assembly schedules," Proc. Sympos. Appl. Math., vol. 10, American Math. Soc., (1960) 281-289.

[28] Gale, D., "A theorem on flows in networks," Pacific Journal of Math. 7 (1957) 1073-1082.

[29] Ghouila-Houri, A., "Un résultat à la notion de diametre," C. R. Acad. Sci. Paris <u>250</u> (1960) 4254-4256.

[30] Gilbert, E. N., "Enumeration of labelled graphs," Canadian Journal of Math. <u>8</u> (1956) 405-411.

[31] Gleason, A. M. and Greenwood, R. E., "Combinatorial relations and chromatic graphs," Canadian Journal of Math. <u>7</u> (1955) 1-7.

[32] Goodman, A. W., "On sets of acquaintances and strangers at any party," American Math. Monthly <u>66</u> (1959) 778-783.

[33] Gumbel, E. J., "Statistics of Extremes," New York, Columbia University Press, 1958.

[34] Harary, F. and Ross, I. C., "On the determination of redundancies in sociometric chains," Psychometrika <u>17</u> (1952) 195-208.

[35] Harary, F. and Norman, R. Z., "The dissimilarity characteristic of Husimi trees," Annals of Mathematics <u>58</u> (1953) 134-141.

[36] Harary, F. and Norman, R. Z., "The dissimilarity characteristic of linear graphs," Proc. American Math. Soc. <u>5</u> (1954) 131-135.

[37] Harary, F. and Ross, I. C., "The number of complete cycles in a communication network," The Journal of Social Psychology <u>40</u> (1954) 329-332.

[38] Harary, F., "The number of linear, directed, rooted, and connected graphs," Trans. American Math. Soc. <u>78</u> (1955) 445-463.

[39] Harary, F., "Note on an enumeration theorem of Davis and Slepian," Michigan Math. Journal 3 (1955-56) 149-153.

[40] Harary, F., "The number of oriented graphs," Michigan Math. Journal 4 (1957) 221-224.

[41] Harary, F., "Note on Carnap's relational asymptotic relative frequencies," Journal of Symbolic Logic 23 (1958) 257-260.

[42] Harary, F., "On the number of bi-colored graphs," Pacific Journal of Math. 8 (1958) 743-755.

[43] Harary, F., "The number of functional digraphs," Mathematische Annalen 138 (1959) 203-210.

[44] Harary, F. and Prins, G., "The number of homeomorphically irreducible trees, and other species," Acta Mathematica 101 (1959) 141-162.

[45] Harary, F. and Norman, R. Z., "Dissimilarity characteristic theorems for graphs," Proc. American Math. Soc. 11 (1960) 332-334.

[46] Harary, F. and Prins, G., "Enumeration of connected bicolored graphs and bichromatic graphs," Canadian Journal of Math. (to appear).

[47] Harris, B., "Probability distributions related to random mappings," Annals of Math. Statistics 31 (1960) 1045-1062.

[48] Hyltén-Cavallius, C., "On a combinatorical problem," Colloquium Mathematicum 6 (1958) 59-65.

[49] Katz, L., "The distribution of the number of isolates in a social group," Annals of Math. Statistics 23 (1952) 271-276.

[50] Katz, L., "Probability of indecomposability of a random mapping function," Annals of Math. Statistics 26 (1955) 512-517.

[51] Katz, L. and Wilson, T., "The variance of the number of mutual choices in sociometry," Psychometrika 21 (1956) 299-304.

[52] Katz, L. and Powell, J., "Probability distributions of random variables associated with a structure of the sample space of sociometric investigations," Annals of Math. Statistics 28 (1957) 442-448.

[53] Kemeny, J. G., Snell, J. L., and Thompson, G. L., "Introduction to Finite Mathematics," Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1957.

[54] Kendall, M. G. and Babington Smith, B., "On the method of paired comparisons," Biometrika 31 (1940) 324-345.

[55] Kendall, M. G., "Rank Correlation Methods," London, Charles Griffin and Co., Lim., 1948.

[56] Kendall, M. G. and Stuart, A., "The Advanced Theory of Statistics (1)," London, Charles Griffin and Co., Lim., 1958.

[57] König, D., "Theorie der Endlichen und Unendlichen Graphen," Leipzig, Akad. Verl. M.B.H., 1936. Reprinted, New York, Chelsea Publishing Co., 1950.

[58] Kövari, T., Sós, V. T., and Turán, P., "On a problem of K. Zarankiewicz," Colloquium Mathematicum 3 (1954) 50-57.

[59] Landau, H. G., "On dominance relations and the structure of animal societies: I. Effect of inherent characteristics," Bull. Math. Biophysics 13 (1951) 1-19.

[60] Landau, H. G., "On dominance relations and the structure of animal societies: II. Some effects of possible social factors," Bull. Math. Biophysics 13 (1951) 245-262.

[61] Landau, H. G., "On dominance relations and the structure of animal societies: III. The condition for a score structure," Bull. Math. Biophysics 15 (1953) 143-148.

[62] Lee, C. Y., "An enumeration problem related to the number of labelled bi-coloured graphs," Canadian Journal of Math. 13 (1961) 217-220.

[63] Listing, J. B., "Der Census räumlicher Complexe oder Verallgemeinerung des Eulerschen Satz von den Polyedern," Göttinger Abhandlungen 10 (1862).

[64] Mohanty, S. G. and Narayana, T. V., "Some properties of compositions of an integer and their application to probability theory and statistics," (abstract), Annals of Math. Statistics 32 (1961) 637.

[65] Moon, J. W. and Moser, L., "Almost all tournaments are irreducible," Canadian Math. Bull. 5 (1962).

[66] Moran, P. A. P., "On the method of paired comparisons," Biometrika 34 (1947) 363-365.

[67] Moser, L., unpublished manuscript.

[68] Moser, L., and Whitney, E. L., "Weighted compositions," Canadian Math. Bull. 4 (1961) 39-43.

[69] Murnaghan, F. D., "The Theory of Group Representations," Baltimore, John Hopkins Press, 1938.

[70] Netto, E., "Lehrbuch der Combinatorik," Leipzig, 1901. Reprinted New York, Chelsea Publishing Co., 1958.

[71] Otter, R., "The number of trees," Annals of Mathematics 49 (1948) 583-599.

[72] Pearson, K., "Tables of the incomplete beta function," London, Cambridge University Press, 1934.

[73] Peterson, W. W., "Error-Correcting Codes," New York, John Wiley and Sons, Inc., 1961.

[74] Pólya, G., "Kombinatorische Anzahlbestimmungen für Gruppen, Graphen, und chemische Verbindungen," Acta Mathematica 68 (1937) 145-254.

[75] Ramsey, F. P., "On a problem of formal logic," Proc. London Math. Soc. (2) 30 (1930) 71-83.

[76] Read, R. C., "The enumeration of locally restricted graphs (I)," Journal London Math. Soc. 34 (1959) 417-436.

[77] Read, R. C., "The enumeration of locally restricted graphs (II)," Journal London Math. Soc. 35 (1960) 344-351.

[78] Read, R. C., "The number of k-coloured graphs on labelled nodes," Canadian Journal of Math. 12 (1960) 409-413.

[79] Read, R. C., "On the number of functional digraphs," Mathematische Annalen 140 (1961) 108-109.

[80] Rédei, L., "Ein kominatorischer Satz," Acta Litterarum ac Scientiarum (Sectio Scientiarum Mathematicarum), Szeged 7 (1934) 39-43.

[81] Rényi, A., "Some remarks on the theory of trees," Publications of the Math. Inst. of the Hungarian Acad. of Sci. 4 (1959) 73-85.

[82] Rieman, I., "Über ein Problem von K. Zarankiewicz," Acta Mathematica Academiae Scientiarum Hungaricae 9 (1958) 269-279.

[83] Riordan, J., "The number of labelled colored and chromatic trees," Acta Mathematica 97 (1957) 211-225.

[84] Riordan, J., "An Introduction to Combinatorial Analysis," New York, John Wiley and Sons, Inc., 1958.

[85] Riordan, J., "The enumeration of trees by height and diameter," IBM Journal of Research and Development 4 (1960) 473-478.

[86] Roy, B., "Sur quelques proprietes des graphes fortement connexes," C. R. Acad. Sci. Paris 249 (1959) 2151-2152.

[87] Ryser, H. J., "Combinatorial properties of matrices of 0's and 1's," Canadian Journal of Math. 9 (1957) 371-377.

[88] Saintë-Lague, A., "Les réseaux," C. R. Acad. Sci. Paris 176 (1923) 747-750.

[89] Sauvé, L., "On chromatic graphs," American Math. Monthly 68 (1961) 107-111.

[90] Scoins, H. I., "The number of trees with nodes of alternate parity," Proc. Cambridge Phil. Soc. 58 (1962) 12-16.

[91] Slepian, D., "On the number of symmetry types of Boolean functions of n variables," Canadian Journal of Math. 5 (1953) 185-193.

[92] Wright, E. M., "Counting coloured graphs," Canadian Journal of Math. 13 (1961) 683-693.

[93] Zarankiewicz, K., Problem 101, Colloquium Mathematicum 2 (1951) 301.